Combining Ion Mobility Mass Spectrometry and Computational Methods to Study Structures of Biomolecules in the Gas Phase

by

Sugyan M. Dixit

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2019

Doctoral Committee:

Professor Brandon T. Ruotolo, Chair
Professor Charles L. Brooks III
Professor Kristina L. Håkansson
Professor Ayyalusamy Ramamoorthy

Sugyan M. Dixit

sugyan@umich.edu

ORCID iD: 0000-0002-6313-7974

Dedication

This thesis is dedicated to all the teachers I have had during my entire life.

Acknowledgements

So many people have left a positive remark during my journey in getting a PhD. First, I would like to thank my advisor, Prof. Brandon Ruotolo. His endless support has always boosted my willingness to work hard during my graduate career. His high standards in terms of scientific rigour and scientific communications have heavily influenced me in becoming a scientist and will continue to impact my career in academics. In the last few years, he has become more like a colleague than just a mentor because of the back and forth in discussions we have during our meetings. As such, I am certain that I will be seeking his guidance and suggestions in the next phase of my career. So, just a very huge thanks to my advisor for being the awesome mentor he is.

Second, I would like to thank my thesis committee members Charles Brooks (Charlie), Kristina Hakansson (Kicki), and Ayyalusamy Ramamoorthy (Rams). They have played an important role in shaping this dissertation. Their intellectual contributions during my candidacy examination and data meeting have made me think broader in pursuing the goals of my experiments. I would especially like to thank Charlie for his thoughtful insights and suggestions in the biophysics area of this dissertation. Without learning the lessons from the thermodynamics class or the CHARMM workshop conducted by Charlie, my thesis would have been way more challenging.

During my graduate career, I have had the opportunity to work with wonderful scientists in the lab. The lab mates that I worked alongside have always sparked interest in many different aspects of the research in the lab. Joseph D. Eschweiler (Joe) introduced me to programming,

specifically python. He was also the only other person in the lab involved in the computational modelling of proteins and protein complexes. So, I always had good discussions with him. Next, I would like to say huge thank you to Daniel Polasky (Dan). While I was learning python and becoming quite good at it, he showed me even better ways of writing efficient and modular programs. The collaborative projects with him have always been joyful. We always had similar goals in our minds, and we tackled different parts of the project to complete it. Not only were we lab mates for the entirety of my graduate school, but we were also flatmates. I would then like to thank the rest of my amazing lab mates: Sarah Fantin (for creating the GUI design for CIUSuite2, organizing social drink hours, bringing delicious baked goods), Cara D'amico (for bringing fun in the lab and getting me involved in the droplet fast CIU project), Chunyi Zhao (for organizing birthdays in the lab, bringing delicious snacks, and working together on some modeling projects), Daniel Vallejo (for always being nice and creating some wicked graphics on illustrator), visiting undergraduate student Hua Pan (for being my best mentee so far and acquiring the challenging nanodisc-peptide complex data on the instrument), Varun Gadkari (for being a great support system and having good discussions on fundamentals of ion mobility), Kristine Parson (for just being herself all the time, bringing joy to the lab, and working together on modeling membrane proteins), Yilin Han (for having conversations about amyloid project and international student issues), Carolina Rojas Ramirez (for believing in top-down mass spectrometry to a whole another level and having good conversations regarding statistics and programming), Chae Kyung Jeon (for being a good mentee and taking over modelling projects), and Brock Juliano (for bringing positive enthusiasm in the lab).

In addition to my coworkers, I have to thank my collaborators from Waters Corporation. I have had a great privilege working with Kevin Giles, Keith Richardson, and David Langridge for

the travelling wave ion mobility (TWIM) section of the thesis. Chapters 3 and 4 were a joint effort of an almost 2 year-long collaboration. I learned new TWIM theory from them and got involved in projects that will result in publication.

A special thanks also goes to my undergraduate research advisor, Prof Daniele Fabris from SUNY Albany. He introduced to me native mass spectrometry techniques. He was also the first mentor to challenge me to think critically for solving problems. During my undergraduate research, Matteo Scalabrin, who was a post doc in the lab, mentored me and gave me freedom to lead the research. I knew I wanted to pursue doctoral studies due to my undergraduate research experience.

My friends are an important part of my life, and, thus, this thesis. Without them, I would not have been able to finish my thesis as smoothly as I did. Bipul Chitrakar, who is one of my good friends in Nepal, has always believed that I'd thrive in the US and become successful. After leaving my friends and family in Nepal, I have made friends here whom I consider a part of my family. Eliza K.C., Sandip Khatri, Sahan Shrestha, and Erick Daly fall in that category and they constantly support me. Finally, Alec Valenta, who is my brother and my best friend, has literally had my back through the ups and downs during the last 5 years in Ann Arbor, Michigan. For that, I'd really like to thank him. So much love to all these friends and I feel very lucky to have them in my life.

Finally, I have to thank my family. My mom and dad have made great sacrifices for me throughout their lives. The reason I could travel to the US for higher education, the reason I graduated from SUNY Albany, and the reason I'm getting my PhD are because of my parents. So, huge thanks to them. I'd also like to thank my sister Pragyee Dixit for always believing in me. I also have to thank the family I have in NY, Rizbi Dixit, Pragyan Dixit, and Shreya Bhattarai. They have always been very supportive and it's always fun to spend time with them.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

Mass spectrometry (MS)

Ion mobility mass spectrometry (IM-MS)

Traveling wave ion mobility (TWIM)

Drift tube ion mobility (DTIM)

Arrival time distributions (ATDs)

Collision cross section (CCS)

Collision induced unfolding (CIU)

Nanodisc (ND)

Molecular dynamics (MD)

Abstract

Characterizing the complex, dynamically regulated networks in cells is critical for the understanding of disease mechanisms and development of therapeutics. Over the last two decades, mass spectrometry (MS) has emerged as a key structural biology tool enabling rapid analysis of complex samples. Native MS has had tremendous success in the structural elucidation of proteins, protein complexes, and protein-ligand interactions. Ion mobility MS (IM-MS), under the native MS category, has gained popularity as a structural biology technique capable of reporting collision cross section (CCS) area of biomolecular ions that can be used as an attribute for identification in bioinformatics workflows and restraint for generating three-dimensional models of proteins. Traveling wave IM (TWIM) is the most used IM platform across research and industry laboratories. However, the amount of information and the accuracy obtained from TWIM measurements have been compromised due to the lack of fundamental understanding of the technology itself. Therefore, in this thesis, novel developments in IM-MS techniques, especially with TWIM, are described that are capable of providing accurate biophysical measurements of proteins and protein complexes in a high throughput manner.

In chapter 2, we devise a semi-empirical relationship that can model TWIM arrival time distributions (ATDs) across a range of TWIM conditions. A conformational broadening parameter can be extracted from the semi-empirical formalism that describes the size of the structural heterogeneity of biomolecules in the gas phase. We validated our method by investigating the origins of structural heterogeneity arising in a set of model peptides. The conformational broadening parameter also properly reflected the reduction in structural flexibility when we

introduced cross-links in a protein complex. In chapter 3, we described a novel pseudo-trapping phenomenon in TW ion guides that produces aberrant ATDs. This was described using a theoretical model and ion trajectory simulations highlighting that imperfect TW leads to a repetitive pattern of ion motion causing the ions with even small mobility difference to travel with the same mean velocity. Consequently, the ions' transit times through the device were altered detrimentally affecting the calibrated CCS values. In chapter 4 we show new calibration functions capable of generating precise and accurate CCS values from TWIM measurements. Velocity relaxation and travelling wave edge effects are incorporated into the new function termed as blend + radial that outperforms the current calibration function in terms of accuracy, preciseness, and robustness. We benchmarked the new function using a large scale of analyte ions comprised of small molecules and metabolites, peptides, denature proteins, and native proteins.

The last chapter showcases the utility of IM-MS platform for high throughput characterization of protein structure and protein-ligand interactions using collision induced unfolding (CIU) experiments. A classification algorithm was built for a single state and multi-state classification of CIU fingerprints, where a state can be defined as charge states of the ions, protein incubation properties, etc. Using our classification workflow, we were able to identify the class of an unknown endogenous lipid in a membrane protein complex. Multi-state classifier boosted the accuracy of the classification model, which was demonstrated using Src-kinase ligand binding experiments and biotherapeutic innovator and biosimilar comparisons.

Overall, the developments in the IM-MS methods, especially the theoretical contributions to TWIM technology, described in this thesis will allow the widespread TWIM community to properly utilize the platform in the areas of chemical analysis and structural biology.

Chapter 1 Introduction

The study of protein structure, function, and complex formation is of great importance for human health. Proteins comprise the machines that are responsible for performing almost every cellular activity. Therefore, an improved understanding of protein function can lead to the elucidation of disease mechanisms, as well as the downstream development of therapeutic stratagies.[1] Protein functions are highly correlated to their structures and dynamics.[2] In the last decade, there have been a plethora of advancements in structural biology techniques aimed at solving structures of proteins and protein complexes.[3–8] However, due to their inherent structural heterogeneity and dynamics,[9] as well as the complex environment[10,11] often associated with native protein complexes, protein structure determination still remains a challenge. Here, I introduce some of the state-of-the-art techniques used to determine protein structures, with emphasis on structural mass spectrometry methods, as these are the main subject of my dissertation.

1.1 Methods for protein structure characterization

The first report of an atomic-resolution protein structure from Kendrew et al[12] in 1958 utilized X-ray diffraction data to analyze myoglobin crystals. In 1971, the Protein Data Bank (PDB) was established to deposit the atomic coordinates for similarly-solved protein structures.[13] Since then, there has been an exponential growth in the number of PDB entries; including structures solved using X-ray, nuclear magnetic resonance (NMR), and electron

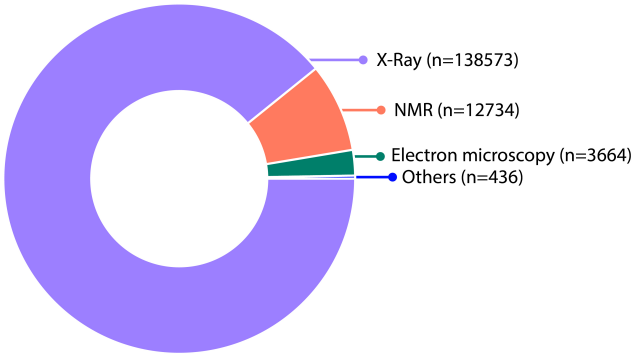


Figure 1-1. PDB data distribution (Year 2019) sorted by experimental methods.

microscopy (EM) methods. Figure 1-1 displays a recent snapshot of structure entries, categorized based on the different methods used to create the entries.

It is important to note that many proteins systems remain refractory to high-resolution structural biology approaches, necessitating lower-resolution and integrative workflows to create models. Such models recently began to be deposited alongside higher-resolution structures in the PDB.

1.1.1 Experimental methods for protein structure determination

Most of the entries in PDB originate from X-ray crystallography (Figure 1-1). This method requires that proteins are purified and crystallized prior to analysis. Crystals are then subjected to intense X-ray beams. By subjecting the crystals to such X-rays, the protein in the crystal diffracts the beam, yielding characteristic patterns that can be analyzed and mapped to produce the distribution of electrons in the crystalized protein. The electron density map can then be used to locate each atom within the protein. Thus, X-ray crystallography can provide protein structures that possess detailed atomic information, including interactions with ligands, ions, and other molecules.[14,15] However, the crystallization process and obtaining adequate diffraction data can be limiting steps in x-ray crystallography, often resulting long and uncertain timelines for protein structure generation.[16,17] Therefore, biochemists and molecular biologists spend a large amount of time and expense in order to express and purify large amounts of homogenous proteins in an effort to form well-ordered crystals. Sometimes this process includes creating mutant or truncated protein forms, as well as including small molecules or ligands that help to stabilize proteins. Therefore, despite producing high-resolution protein structures, the structural throughput of X-ray diffraction remains a challenge.[18–21]

NMR spectroscopy can be used to determine atomic-level structures of proteins. In NMR, purified proteins are placed in a strong magnetic field and exposed to radio frequency excitation pulses in an effort to measure nuclear spin resonance. The observed resonances can be used to characterize of the local bonding environment of nuclei, ultimately generating a list of restraints that can be used to build three-dimensional atomic models of proteins.[22–25] An advantage of NMR over X-ray in the context of protein structure determination is that NMR can directly probe protein conformers in solution. As such, NMR analyses yield an ensemble of protein structures that are consistent with the experimentally observed restraints.[26–28] Recent efforts have sought to extend NMR spectroscopy to the study larger proteins, membrane proteins, and multi-protein complexes.[28–33] One challenge associated NMR techniques, as is also the case with X-ray crystallography, is their requirement of large amounts of purified proteins.[34]

Electron microscopies (EM), especially cryo-EM variants, represent a dynamic emerging group of technologies for high-resolution protein structure analysis. In EM, the scattering of an electron beam is used to evaluate thousands of protein-based particles housed either on a grid, or within vitrified ice (for cryo-EM). Particle images are grouped into classes, and a three-dimensional electron density map can be constructed through a detailed analysis of many such two-dimensional images, assuming that each represents particles of identical composition and structure.[35] EM methods have yielded structures at resolution values comparable to X-ray diffraction.[3,36–38] Furthermore, EM has specifically been utilized in integrative modeling methods to generate structures of large macromolecular assemblies such as ribosomal complexes[39,40] and membrane proteins.[41–43] While EM techniques do not typically require the large amounts of sample often consumed by comparable X-ray and NMR methods, cryo-EM optimization does require highly-purified samples and significant efforts in buffer optimization to produce high-quality data.

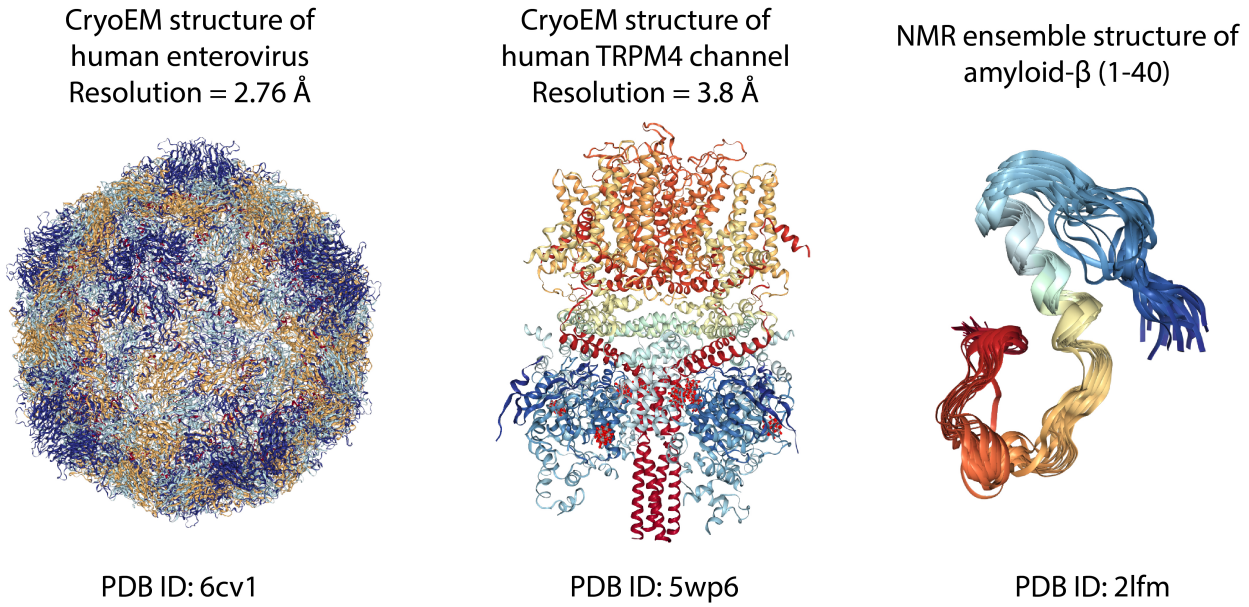


Figure 1-2. Atomic level structures of proteins and protein complexes determined via EM and NMR.

Even though the techniques listed above are capable of generating high resolution protein structures (Figure 1-2), such methods are not always practical due to the required amounts of time and sample, cost, and the other limitations listed above. Therefore, integrative structural biology approaches, that seek to combine data acquired from lower-resolution structural techniques, have emerged to assess the structures of proteins and complexes that remain refractory to NMR, X-ray and EM techniques. Integrative structural biology workflows can include data from techniques such as small angle X-ray scattering (SAXS)[44,45], circular dichroism (CD)[46,47], and Fourier transform infrared spectroscopy (FTIR)[6,48]. The techniques included in integrative structural campaigns vary, but most are included due to their ability to extract valuable restraint information from relatively small amounts of unpurified protein. Future development of high throughput protein structure characterization techniques is clearly required in order to fill our current protein structure knowledge gap, especially given the growing scopes of metagenomics, proteomics, and bioinformatics approaches.

1.1.2 Computational Methods

From fitting atomic models to experimental observations or generating three-dimensional structure via *ab initio* or homology modeling, the use of computational methods in protein structure characterization has always been critical. Despite decades of effort, predicting protein structure from its amino acid sequence remains challenging.[49] Both *ab initio* and template-based modeling have been used to predict accurate protein structures.[50,51] Template-based and integrative modeling has been vital for structure determination of macromolecular complexes.[52,53] In this section, selected computational methods used in protein structure characterization are summarized.

Ab initio modeling approaches seek to perform a search of protein conformational space using a designed, and thoroughly evaluated, energy functions. This process usually generates many thousands of possible conformations from which final protein models are selected. Three factors typically dictate the accuracy of protein structure predictions from such *ab initio* modeling approaches. First, the accuracy of the energy function used to perform the molecular dynamics (MD) simulation ultimately determines the ability of such tools to differentiate between the most thermodynamically stable (native) state and structural decoys for a given protein sequence. There are several energy functions that are based on physical forces, chemical knowledge, or a combination of both.[54–59] Below is an example of physics based CHARMM[60] energy function.

$$E_{bonded} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{\substack{improper \\ dihedrals}} K_\varphi(\varphi - \varphi_0)^2 + \sum_{dihedrals} \sum_{n=1}^{6} K_{\phi,n}\,(1 + \cos(n\phi - \delta_n)) \qquad (1-1a)$$

$$E_{nonbonded} = \sum_{\substack{nonbonded \\ pairs\ i,j}} \frac{q_i q_j}{4\pi D \|r_i - r_j\|} + \sum_{\substack{nonbonded \\ pairs\ i,j}} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{\|r_i - r_j\|} \right)^{12} - 2 \left(\frac{R_{min,ij}}{\|r_i - r_j\|} \right)^{6} \right] \qquad (1-1b)$$

The CHARMM force field is the additive potential energy function from the bonded and nonbonded terms. The bonded elements displayed in equation 1-1a consist of terms that correspond to various bond lengths and bond angles within a molecule. The nonbonded potential energy function shown in 1-1b comprises terms describing electrostatic and van der Waals terms. ROSETTA[59], I-TASSER[61–63], and QUARK[64,65] are among the methods that use a combination of knowledge and physics-based energy terms for *ab initio* structural prediction. In ROSETTA, for instance, fragments are assembled from unrelated protein structures found within the PDB to build full length models using a knowledge-based force field. Such models are then refined using a physics-based MD approach. Subsequently, an efficient conformational search algorithm helps generate low energy structures. Widely used algorithms for physics-based MD include: annealing Monte Carlo (MC), replica exchange MC method (REM), simulated annealing molecular dynamics (MD), replica exchange MD (REMD), accelerated MD, and genetic algorithm (GA).[50,66] Finally, a good selection strategy aids in filtering near native like structures from the pool of decoy structures. This usually includes clustering of low energy structures based on their structural similarity.[50] On-going efforts to improve physics and knowledge-based force fields as means for protein structure prediction are matched by efforts to include experimental data in an effort to guide protein structure prediction workflows. For example, the Integrative Modeling Platform (IMP) provides an environment for comparative protein structure modeling that enables data-based restraints that has been successful in creating large macromolecular complexes such as nucleopore protein complex.[53,67] The growth of such integrated methods clearly necessitates the continual development of structural biology measurement tools that seek to increase the scope of accessible proteins and complexes for structure determination.

1.2 Mass Spectrometry in structural biology

Over the last two decades mass spectrometry (MS) has emerged as a key approach in structural biology[68–72]. A pyramid schematic of different MS methods used in structural biology is shown in Figure 1-3. MS has generally been associated with bottom-up proteomics which is used for 1) identifying proteins including any modifications such as post -translational modifications (PTMs), 2) measuring changes in protein expression levels, and 3) investigating protein-protein interactions.[73] Similar goals can be achieved using middle-down[74–77] and top-down proteomics[78–80]. Oxidative foot-printing[81–83], hydrogen deuterium exchange[84–88], and cross-linking[89–93] methods can be combined with these MS workflows to obtain more detailed structural information from proteins and protein complexes. Native MS provides the means to directly probe the structure and dynamics of protein and protein complexes in the gas phase.[94–98] In native MS ions are transferred to the gas-phase using nano-electrospray ionization (nESI) using a non-denaturing solution, thus retaining their non-covalent interactions.[99–104] The mass resolution provided by contemporary mass analyzers has made it possible to determine protein oligomeric states and the binding of ligands. In tandem with native MS, fragmentation techniques, such as collision induced dissociation (CID)[105–108], surface induced dissociation (SID)[109–112], electron transfer dissociation (ETD)[113,114],

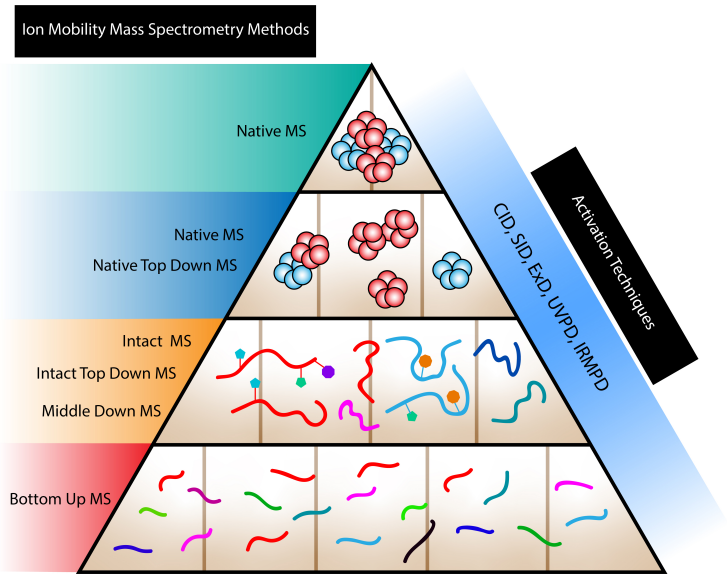


Figure 1-3. Different structural MS methods to study proteins and protein complexes.

electron capture dissociation (ECD)[115–118], and ultraviolet photo dissociation (UVPD)[119–122], can provide information ranging from protein primary to quaternary structure. Native MS is emerging as a powerful structural biology tool that can be used to study the structures of unpurified proteins and protein complexes using a limited amount of sample.

1.2.1 Native Ion Mobility – Mass Spectrometry

Ion mobility – mass spectrometry (IM – MS) is a structural MS method that uses IM to determine to the orientationally averaged collision cross section (CCS) of protein ions in the gas phase.[123–126] During IM separation, protein ions traverse through a gas filled chamber under the influence of a weak electric field (E) and are separated according to their mobility (K). In drift tube IM (DTIM)[127] the velocity of the ions (v_{ion}) can be directly related to their mobility by using equation 1-2

$$v_{ion} = KE \qquad 1-2$$

In the low-field limit, where the measured mobility is independent of the drift field, K is given by Mason-Shamp's relationship[128] as shown below

$$K = \frac{(18\pi)^{\frac{1}{2}}}{16}\frac{ze}{(k_bT)^{\frac{1}{2}}}\left(\frac{1}{m_I}+\frac{1}{m_N}\right)^{\frac{1}{2}}\frac{1}{\Omega}\frac{1}{N} \qquad 1-3$$

where k_b is the Boltzmann constant, z is the ion charge, e is the elementary charge, m_I is the mass of the ion, m_N is the mass of the neutral gas, N is the neutral gas number density, and Ω is the orientationally averaged CCS. Ion transit time (t_d), which is related to the ion velocity, can be used to infer CCS. With a nested IM – MS experiment, we gain three-dimensional information comprising of m/z ratio, arrival time (t_d) distributions (ATD), and intensity of ions, as shown in

Figure 1-4. Therefore, in a native IM – MS experiment protein's oligomeric as well as conformational states can be distinguished.

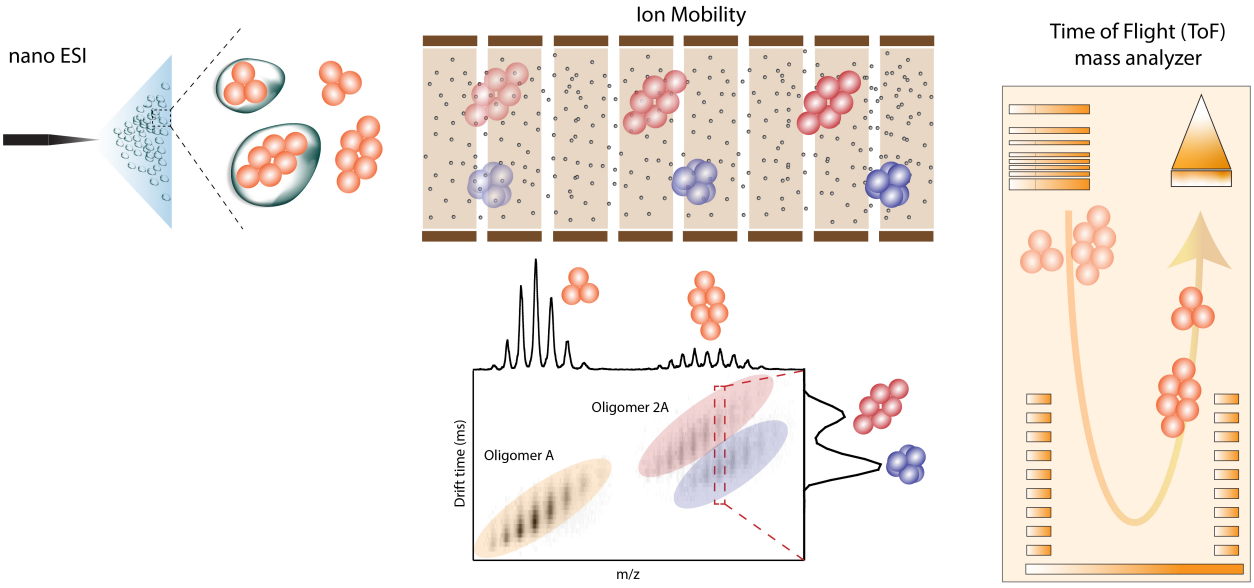


Figure 1-4. Schematic of IM-MS experiment. Ions generated via nESI are separated by IM and analyzed by ToF. The m/z, drift time, and intensity value contour plot shows two different oligomeric states as well as different conformational states in Oligomer 2A

IM-MS has been successfully applied to gain structural information on many proteins and protein complexes. IM – MS analysis of amyloid forming entities such as amyloid-β, α-synuclein, and β_2-microglobulin has shown not only multiple oligomeric states but also multiple structural populations within each oligomeric state.[129–134] In addition, CCSs obtained from IM measurement have been used to filter structures from MD simulations[135,136] by calculating theoretical CCSs[137–139] of models. CCS is also used as a restraint for generating coarse grain topology models which can then be merged with atomic models.[140,141] Using this strategy, IM-MS integrative modeling has been used to model systems such as tryptophan synthase[141], nitrobenzene dioxygenase[142], the DNA clamp loader complexed including single stranded DNA binding protein[143], the Yeast eIF3:eIF5 complex[144], toyocamycin nitrile hydratase[109], and the urease activation complex[145]. A recent computational study [146] revealed that significant ambiguities exist when defining multi-protein models using subunit-level CCS values, which can be significantly reduced by moving to

sources of domain-level information to generate such topology models. Thus, an aim of this thesis is to reduce such ambiguities in models by rapidly obtaining information on protein structure dynamics and by determining CCS values more accurately.

1.2.2 Traveling wave ion mobility

Prior to the introduction of traveling wave IM (TWIM) technology in 2004[147] IM experiments were mostly conducted using home-built DTIM analyzers. Since the commercialization of TWIM coupled to MS in a quadrupole/TWIM/orthogonally accelerated - ToF instrument platform starting around 2006[148], IM-MS has been brought to a much wider audience spanning many laboratories in both academia and industry[149]. Subsequently, TWIM has been implemented within structures for lossless ion manipulation (SLIM) devices, enabling very high (R ≈ 340, a 5 fold improvement over typical DTIM devices)[150]. A TWIM device is comprised of a radio frequency (RF) ion guide with repeating DC pulses applied to a pair or two pairs of ring electrodes sequentially along the device creating a traveling wave (TW) of defined wave amplitude and velocity (Figure 1-5).[151–153] Lower mobility ions are overtaken by the waves, resulting in a "roll-over" event, which occurs less often for higher mobility ions. Thus, higher mobility ions travel faster than those having lower mobility, enabling mobility-based separations.

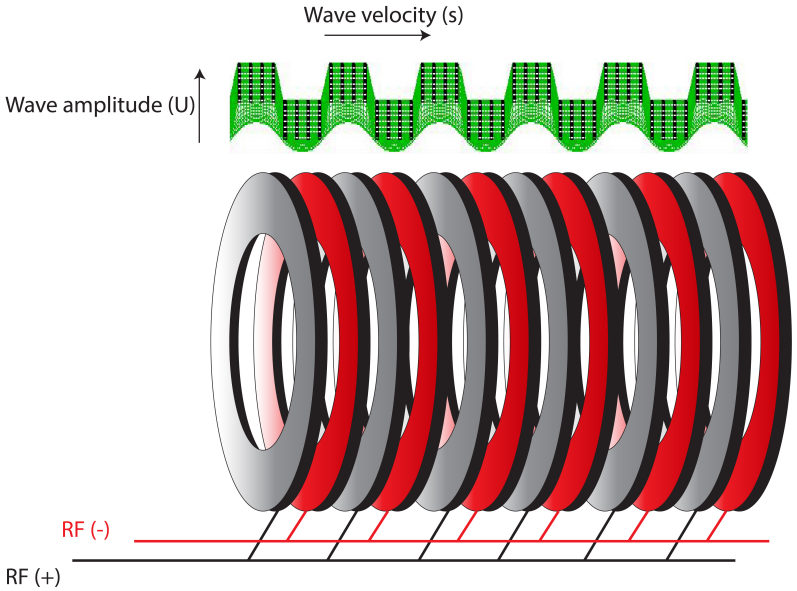


Figure 1-5. Schematic of TWIM analyzer. TWIM analyzer comprised of RF applied stacked ring ion guide with DC pulses applied in pair to create a traveling wave of wave amplitude U and wave velocity s.

Despite being widely used, there are a number of challenges associated with the continued development of TWIM. For example, prior to work described in this thesis, it was not possible to predict TWIM ATDs due to deficiencies in TWIM theory. In contrast, DTIM ATDs can be used to directly quantify the structural heterogeneity of ions.[127,154] Furthermore, calibration of TWIM arrival times is required to obtain CCS values from the technology, whereas DTIM measurements can be used to obtain such information without the need for standards.[155] As above, this gap results directly from a lack understanding surrounding ion transport in TWIM. In 2008, Shvartsburg et al[152] first derived a relationship describing ion velocity in TWIM using equation 1-4.

$$v_{ion} = K^2 \frac{\int_0^{\lambda} \frac{E^2(X)dX}{v^2 - K^2E^2(X)}}{\int_0^{\lambda} \frac{vdX}{v^2 - K^2E^2(X)}} \qquad 1-4$$

where v_{ion} is the velocity of the ion, K is ion mobility, X is the position of ion in X direction in TW reference frame, E is the electric field strength, and v is the TW velocity. This study generated equation 1-6 to define the TWIM resolution, which is defined by equation 1-5, as:

$$R = \frac{t_d}{\Delta t_d} \qquad 1-5$$

$$R = \frac{1}{4}\left(\frac{qKE^2L}{vk_bTln2}\right)^{\frac{1}{2}} \qquad 1-6$$

where t_d and Δt_d are centroid and full width half max of TWIM ATDs, respectively, q is the charge of the ion, and L is the length of the TWIM cell. These models, however, do not faithfully reproduce experimental TWIM, as shown in Figure 1-6. In 2017, Mortensen et al[156] developed a computational method to calibrate TWIM drift times to obtain CCS values. Although promising, this method exhibits poor prediction accuracies for certain ions and TWIM conditions. Finally,

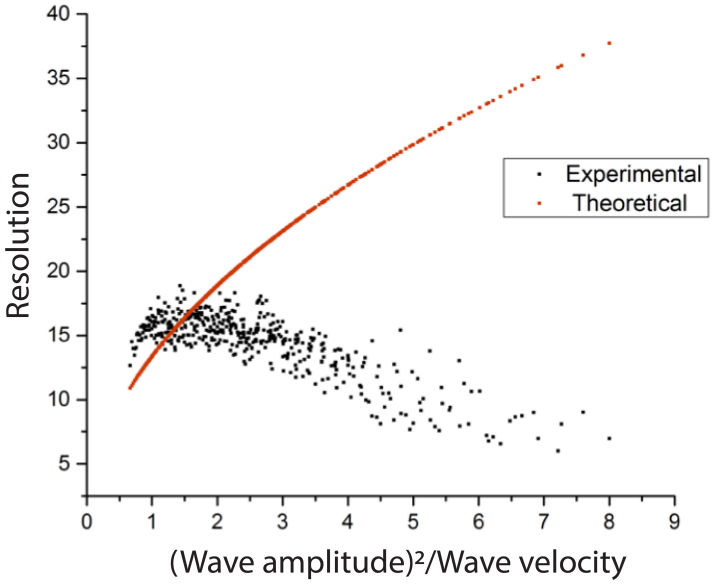


Figure 1-6. Comparison of the experimentally observed and calculated resolution (using equation 1-6) in across a range of TWIM wave amplitude and velocity conditions.

Richardson et al[153] revisited the fundamentals of TWIM generating a series of relationships describing ion motion by incorporating velocity relaxation for ions as well as the anharmonicity present in practically realized TWIM. Equation 1-7 describes the ion motion in TW reference frame

$$\alpha\ddot{z} = -\dot{z} - 1 + \gamma sinz \qquad 1-7$$

where

$$z = kx - \omega t + \delta \qquad 1-8$$

$$\alpha = \omega\frac{Km}{q} \qquad 1-9$$

$$\gamma = \frac{kKV_0}{v} \qquad 1-10$$

where k is the wave number, δ is the phase shift in TW frame, ω is the angular frequency, α relates to the ion's relaxation time, K is ion mobility, V_0 is the amplitude of the TW potential, v is TW velocity, and q is the charge of the ion. The authors incorporated these ideas and proposed new CCS calibration functions in TWIM[155]. Since TWIM is currently the most-widely deployed form

of IM-MS instrument platform, there is a clear need for improved theoretical frameworks that enable accurate CCS determinations. This thesis, in part, accommodates this need through advancements in theory to better understand TWIM ATDs and determine accurate CCSs.

1.2.3 Collision induced unfolding

As discussed in previous sections, activation techniques have been long used to probe the structure and stability of protein ions in the gas phase. Collision induced unfolding (CIU) is an extension of this where collisional activation is followed by IM-MS to probe the conformations of

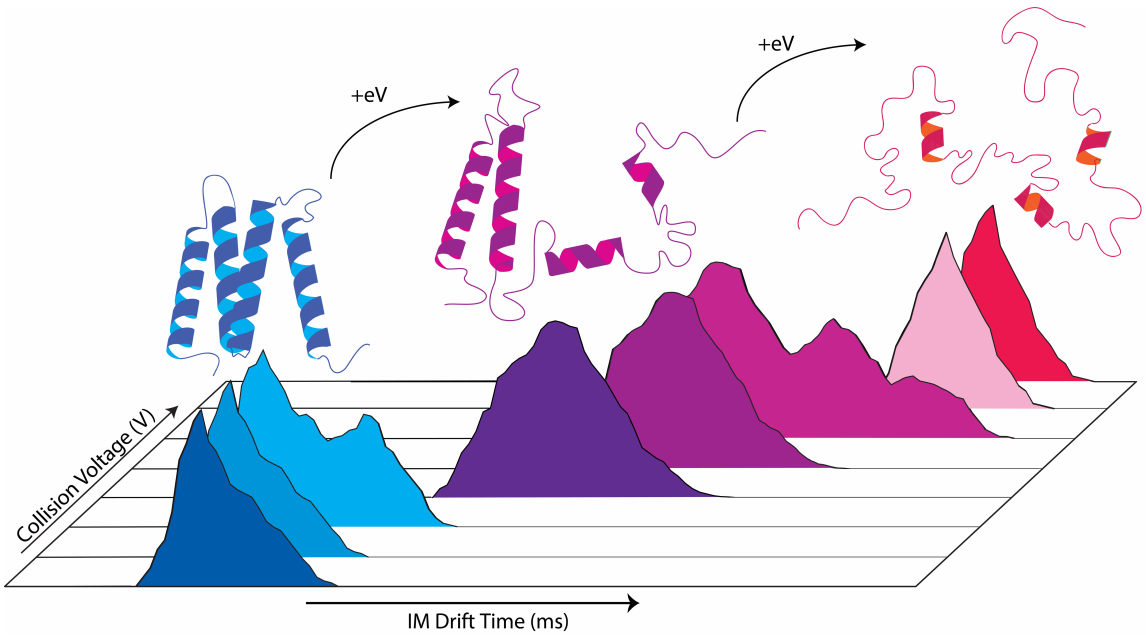


Figure 1-7. Collision induced unfolding of a protein ion. Increasing collision activation causes protein ions to unfold which is monitored using ion mobility arrival time distributions.

protein ions the gas phase, as shown in Figure 1-7.[157] Early CIU experiments probed the activation energy barriers associated with gas-phase folding and unfolding of apo myoglobin following charge manipulation, revealing clear evidence of both Coulombic and structural components for the barriers detected.[158] Since then, CIU has been implemented to study the structure and dissociation behavior of various other proteins and protein complexes.[159,160,169,161–168] This technique has also been extensively used as a fingerprinting method to differentiate and/or classify protein systems. For example, studies by Tian et al[164,170] have showed how CIU can be used to

characterize monoclonal antibodies. Therefore, native IM – MS along with CIU experiments have proven useful in the rapid analysis of protein stability and structure.

1.3 Summary

This dissertation presents efforts to install new and improved capabilities for native IM – MS in the study of biomolecular structure generally. Specifically, this work can be divided into two broad areas: 1) efforts to advance TWIM ion transport theory, and 2) efforts to develop new classification algorithms for CIU.

Chapters 2, 3, and 4 focus on developing different aspects of TWIM theory. In chapter 2 a semi-empirical relationship was derived to predict the resolution and width of TWIM ATDs in order to obtain information on structural heterogeneity of protein ions rapidly using IM-MS. This work has been published as **Dixit, S. M. and Ruotolo, B. T. A Semi-Empirical Framework for Interpreting Traveling Wave Ion Mobility Arrival Time Distributions. *J. Am. Soc. Mass Spectrom*. 2019, *30* (6), 956-966**. Chapter 3 reports the first observation, and theoretical description, of a novel pseudo-trapping of ions within TWIM ion guides. This form of ion trapping can serve as a complicating factor in accurately interpreting TWIM ATDs and obtaining CCS values. Finally, chapter 4 presents and evaluates a new TWIM CCS calibration method, which leverages a previous theoretical treatment of TWIM ion transport and a large TWIM dataset in order to produce a validated TWIM calibration method exhibiting substantially enhanced robustness, accuracy and precision relative to previous method.

Finally, this dissertation ends with chapter 5 that discusses new classification algorithms built to differentiate CIU datasets. A robust statistical method that uses machine learning tools to generate supervised classification workflow that can handle any number of classes and datasets was constructed an implemented within the CIUSuite2 software tool. Furthermore, a multiplexed

classification protocol was generated to pool data acquired across multiple charge states or stress states in order to enhance classification accuracy and prediction power. Development of these classification algorithms has been previously published in the following papers: 1) **Polasky, D. A., Dixit, S. M., Fantin, S. M., and Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data.** ***Anal. Chem.*** **2019, *91* (4), 3147-3155** and 2) **Polasky, D. A., Dixit, S. M., Vallejo, D. D., Kulju, K. D., and Ruotolo, B. T. An Algorithm for Building Multi-State Classifiers Based on Collision-Induced Unfolding Data.** ***Anal. Chem.*** **2019, *91* (6), 10407-10412**.

Chapter 2 A semi-empirical framework for interpreting traveling wave ion mobility arrival time distributions

(Adapted with permission from: Sugyan M. Dixit and Brandon T. Ruotolo. A Semi-Empirical Framework for Interpreting Traveling Wave Ion Mobility Arrival Time Distributions. *J. Am. Soc. Mass. Spectrom.* **2018**, *30* (6), 956-966)

2.1 Introduction

The functions of biomolecules are inherently linked to their structures, making their study critical for wide-ranging research efforts in biochemistry and human disease[1]. Ultimately, protein functions depend upon discrete motions on the atomic scale, giving rise to structural ensembles that are responsible for carrying out various cellular processes[2]. As such, the quantitative assessment of structural ensembles is critically important in understanding the mechanistic details of biomolecular function. For example, proteins adapt their structure to many different binding partners, and therefore, can exhibit large conformational heterogeneity[171]. In addition, a significant fraction of the proteins in eukaryotes contain disordered regions that are involved in many important biophysical processes, but are currently insufficiently understood[172]. The prevalence of dynamics and flexibility in our understanding of protein biophysics has stimulated the development of many novel analytical techniques and computational modeling tools aimed at the detailed assessment of proteins in motion[173].

Despite the importance of protein dynamics, its quantification and characterization has remained a challenge for biophysical measurement techniques for decades. Nuclear magnetic resonance (NMR) spectroscopy is able to probe protein dynamics on timescales ranging from

nanoseconds to milliseconds with atomic resolution[174]. In addition, small angle X-ray scattering (SAXS) measurements have more recently begun to provide information on protein dynamics[175]. In conjunction with experimental data, ensembles of structures have been generated computationally in order to study structural microstates and functional disorder in proteins[176,177]. Despite technical advancements, however, the experimental techniques described above require pure, monodisperse, high concentration samples, which severely limit the biomolecular ensembles that can be probed. Furthermore, the computational sampling of protein dynamics remains challenging due to the difficulties in extending simulation times to match those relevant for most biological processes and a general inability to completely account for configurational entropy in such simulations[178].

By virtue of soft ionization techniques that enable the introduction of solvent-free biomolecular structures in the gas phase[179], mass spectrometry (MS) methods such as native MS[94,180,181], tandem MS in combination with ion activation methods[182–184], hydrogen deuterium exchange (HDX) MS[185,186], and chemical cross-linking (CXL) MS[93,187] have been used to study the structure of proteins and protein complexes using small amounts of sample[188]. Ion mobility (IM)-MS is a structural MS method currently undergoing a period of rapid development, capable of separating protein ions according to their orientationally-averaged size on the millisecond timescale[189]. Several types of IM separators have been coupled to MS, each having their own strengths and weaknesses[125,190]. For instance, drift tube IM (DTIM) works by introducing a time-defined packet of ions into a chamber containing both a weak electric field (E) and an inert buffer gas[191]. Under typical DTIM conditions, ion arrival times are directly proportional to their orientationally-averaged collision cross sections (CCSs), values which serve as coarse-grained structural restraints in biomolecular modeling efforts[192,193]. DTIM arrival time distributions

(ATDs) are well characterized by theory, and have previously been used to assess the structural heterogeneity of biomolecules in the gas phase[154,158,194–196]. Characterization of such heterogeneity reports on the ensemble of structures in solution as well as dynamics in the gas phase. Despite this, traveling wave ion mobility (TWIM)[148,151,197], which uses time-varying electric fields to achieve IM separations, is the most prevalent form of IM-MS for work in the area of structural biology[190]. In contrast to DTIM, assessments of TWIM peak widths in an effort to elucidate biomolecular dynamics and structural heterogeneity, as has been done previously in DTIM measurements, remains challenging due to our currently incomplete understanding of ion transport within TWIM analyzers [152,153,198,199].

In this report, we construct a semi-empirical model that describes the widths of TWIM ATDs across a wide range of TWIM parameters, and is capable of predicting ATD peak widths for nominally mono-conformational biomolecules for use in assessing the dynamics of biomolecules generally. By using a group of model peptides known to have rigid structures in the gas phase, we tested our empirical expression and detected a subtle structural transition in the peptides as a function of sequence length. Furthermore, we correlated our IM-MS measurements with molecular dynamics (MD) simulations, observing strong correlations between computed structural ensembles and our experimental TWIM ATDs. Finally, we applied our model to analyze the TWIM ATDs of unmodified and cross-linked Avidin ions detecting shifts in protein complex peak widths in a manner correlated with the attachment of CXL agents. We conclude by projecting the general utility of quantitative TWIM peak width assessments for gas-phase structural biology.

2.2 Experimental Methods

2.2.1 Chemical Materials

Ac-Ala_n-Lys peptides were custom made from Anaspec, CA. The peptides were dissolved in 90% TFA (Fisher Scientific, O4901) in water to achieve a final concentration of 1 mg/ml for nano-ESI analysis. Avidin (Sigma-Aldrich, A9275) was prepared in 200mM ammonium acetate at a final concentration of 50 μM (Sigma-Aldrich, 09689). BS3 (Thermo Scientific Pierce, PI-21580) was freshly prepared in HEPES (Sigma-Aldrich, H3375) at pH 7.3 before CXL experiments. DL-polyalanine (Sigma-Aldrich, P9003) was used at a final concentration of 1mg/ml in 49.5/49.5/1 ratio of water/acetonitrile (Fisher Scientific, A9984)/acetic acid (Fisher Scientific, A38212).

2.2.2 Chemical Cross-linking

Avidin samples were buffer exchanged into 200mM HEPES, pH 7.3, using micro biospin 6 columns (Bio-rad, CA). BS3 was added to these samples in the following protein to cross-linking reagent ratios: 1:1, 1:50, 1:150, 1:500, 1:1000, 1:1500, and 1:2000. After incubating the samples for 30 minutes at room temperature, the reaction was quenched by buffer exchanging the cross-linked samples into 200mM ammonium acetate.

2.2.3 TWIM-MS

All data was collected on a Synapt G2 TWIM-MS instrument (Waters, Milford MA). Instrumentation details can be found elsewhere[155]. Briefly, ions were generated using nano ESI, and then pulsed into the TWIM cell with 100μs gate pulse width. The TWIM cell is comprised of a stacked ring ion guide (SRIG), where direct current (DC) voltage is applied to two pairs of ring electrodes in a repeating pattern throughout the cell. A series of DC pulses generates a time varying

potential, defined by its wave amplitude (**V**) and wave velocity (***v***). The ultimate structure of the resultant waveform is nominally sinusoidal [148,151,152,197]. In our experiments, the TWIM separator was operated both at a pressure of 3.5 mbar (200 ml/min and 90 ml/min flow rates for He and N_2, respectively) and 4 mbar (200 ml/min and 100 ml/min flow rates for He and N_2, respectively), for peptide and Avidin ion separations, respectively. TWIM-MS data was acquired at values of **V** ranging from 20 V to 40 V in 2 V increments and values of ***v*** ranging from 200 m/s to 500 m/s in 20 m/s increments. After the TWIM cell, ions travel to a transfer region that transports the mobility separated ions into the orthogonal acceleration (oa) time of flight mass analyzer (ToF). TWIM ATD is recorded by synchronization of the oa-ToF acquisition with the gated release of ions from trap into the TWIM cell. CCS was calibrated using D,L polyalanine peptides at a ***v*** value of 520 m/s and **V** values of 20, 25, and 30 V. He CCS values were used to construct calibration function, which was then used to predict He CCS values in our experiments. Three replicate measurements taken at different days were used for calibration in order to obtain uncertainty values as described previously[155].

2.2.4 SIMION Modeling

A Synapt G2 TWIM cell was simulated using SIMION 8.1 (Scientific Instrument Services Inc., Ringoes, NJ, USA)[200]. Potential array (PA) files were created in order to apply potential to 4 electrodes at once and create a 4 repeat pattern that mirrors the implementation of traveling wave dc voltage in TWIM cell[151,197]. A specific script was written to obtain the voltage and electric field strength in the G2 model for downstream analysis.

2.2.5 MD Simulations

We carried out MD simulations to generate a range of configurations that can be used to decipher structural heterogeneity for each peptide. As we have already described the challenges of

defining structural landscape accurately for biological molecules using MD, the simulations in this study are employed to understand the origins of dynamics in a qualitative manner. MD simulations were performed with CHARMM on a workstation with an Intel Xeon processor with eight CPU cores at 2.50 GHz. CHARMM22 force field was employed as it contains the CMAP correction for improved treatment of peptide backbones to achieve more accurate peptide conformations[201]. Helical Ac-Ala_n-Lys peptides with n from 6 to 19 were constructed in CHARMM by fixing the phi and psi angles to -47 and -67 degrees, respectively, and placing the charge on the Lys residue. Peptides were energy minimized using a 10 step conjugate gradient, followed by 100 steps of an adopted basis Newton Raphson (ABNR) method minimization *in vacuo*. The models were then equilibrated at 300 K for 50 ps, after which, they were subjected to a simulated annealing (SA) cycle. Briefly, the system was heated from 300 K to 1000 K in 10 K increments and cooled to 0 K in 10 K decrements, with each temperature step lasting for a minimum of 100 ps. The system was then equilibrated at 0 K for 500 ps. The lowest energy structure from the SA run was then subjected to gradual heating with final temperature of 300 K, 400 K, and 500 K 100 ns of constant temperature MD simulation was performed at those temperatures saving coordinates every 5 ps, generating 20,000 structures. This was performed three times with different random initial velocities. All CHARMM input script files were written in house.

2.2.6 Hierarchical Clustering

A hierarchical clustering method[202,203] from scipy[204,205] was used to classify structural families extracted from MD simulations. For each constant temperature run, 1000 structures were selected at regular intervals in 100 ns runs for classification. Pairwise RMSD values were calculated for all combinations of structures using an in-house script. Pairwise Euclidean distance

matrix was generated using RMSD matrix using SciPy. The resulting distance matrix was then used for hierarchical clustering using average method.

2.2.7 Theoretical CCS Calculations

IMPACT[137] and IMOS[138,139] were used for CCS calculations for model structures. IMPACT was used for CCS calculations on all the structures resulting from MD simulations. The IMOS diffusive trajectory method, which accounts for diffuse scattering in momentum transfer calculation, was used with He gas at 300 K with 50,000 total gas molecules. Non-integer partial charges were included in the structure from CHARMM. Overall, IMOS trajectory method calculations were employed on 280 total structures from our 300K constant temperature run, with 20 structures extracted from each peptide system, in order the test and validate the IMPACT results obtained.

2.2.8 Data Analysis

TWIM ATD data was extracted using TWIMExtract[206]. For CHARMM output trajectory files, in house scripts were written to analyze the trajectories and extract the structures. IM-MS 3D plots were generated using Driftscope (Waters, Milford MA). Data was analyzed using python, numpy, and scipy[204,205]. Matplotlib[207] was used to generate all the output plots shown in this work.

2.3 Theory: A Semi-empirical Width and Resolution Expression for TWIM

For DTIM separations, Equation 2-1 is an analytical solution to the general transport expression that describes the ATD of a single conformation ion species:

$$\boldsymbol{F(z,t) = C\frac{1}{4(\pi Dt)^{\frac{1}{2}}}\left(v_d + \frac{z}{t}\right)\left[1 - \exp\left(-\frac{r_0^2}{4Dt}\right)\right] exp\left[-\frac{(z-v_d t)^2}{4Dt}\right]} \qquad 2-1$$

ere ***F(z,t)*** is the function estimating the ATD of an ion, D is diffusion constant, t is arrival time, v_d is drift velocity, z is the position of the ion in traverse direction, r_0 is the radius of the drift tube

entrance aperture, and C is a constant that is dependent on the initial formation of ion packet[208]. As discussed above, a similar expression is not currently available for TWIM ATD analyses [152].

Generally, we can define the widths of IM ATDs as a sum of a series of band broadening terms:

$$\boldsymbol{W = W_D + W_P + W_{SC} + W_{RXN} + W_C} \qquad 2-2$$

where $\boldsymbol{W}$ is the width of an IM ATD, and each subscript in Equation 2 indicates the origin of the broadening factor indicated, where $\boldsymbol{D}$ is diffusion, $\boldsymbol{P}$ is gate pulse width, $\boldsymbol{SC}$ is space charge, $\boldsymbol{RXN}$ is reaction chemistry, and $\boldsymbol{C}$ is conformational heterogeneity exhibited in the timescale of IM separation[209–211]. $\boldsymbol{W_{SC}}$ and $\boldsymbol{W_{RXN}}$ typically have a negligible impact on $\boldsymbol{W}$, as ion number densities are kept low and inert neutrals are used for IM separation. As such, practical estimates of $\boldsymbol{W}$ depend only on ion diffusion, pulse width, and conformational heterogeneity. As expressed previously for DTIM separations, we model TWIM peak widths using[209]:

$$\boldsymbol{w^2 = \gamma + \beta t_g{}^2 + \delta t_{diff}{}^2} \qquad 2-3$$

where $\boldsymbol{w}$ is the experimental TWIM ATD width (full width at half the maximum peak height, or $\boldsymbol{fwhm}$), $\boldsymbol{t_g}$ is gate pulse width, $\boldsymbol{t_{diff}}$ is diffusion limited width, and we define $\boldsymbol{\delta}$ as a parameter that describes any non-diffusion broadening incorporated into the total TWIM peak width model to achieve a good fit, referred to below as the conformational broadening parameter. By substituting an expression for diffusion-limited TWIM peak width as defined previously[152] into Equation 2-3, we obtain:

$$\boldsymbol{w^2 = \gamma + \beta t_g{}^2 + \delta \frac{16kTvln2}{qLKE^2} t^2} \qquad 2-4$$

where k is Boltzmann's constant, T is the temperature, v is wave velocity, q is the charge of the ion, L is the length of the TWIM cell, K is ion mobility, E is electric field, and t is arrival time of the ion. By simplifying Equation 2-4 we arrive at:

$$w^2 = \gamma + \beta {t_g}^2 + \alpha \frac{v}{KE^2} t^2 \qquad 2-5$$

where

$$\alpha = \delta \frac{16kTln2}{qL} \qquad 2-6$$

In Equations 2-5 and 2-6, α, β, and γ are fitting parameters, where α and ($\gamma + \beta\ t_g^2$) are the slope and intercept, respectively, obtained from a linear regression between w^2 and vt^2/KE^2. Similarly, β is the slope obtained from the linear regression between w^2 and t_g^2 as previously determined in DTIM systems[209]. In contrast to DTIM however, a linear relationship does not exist between w^2 and t_g^2(Figure I-1), and instead exhibit a complex relationship that is also dependent on other factors such as V and v. In order to simplify downstream expressions of TWIM peak width, we set β equal to 1 in the analyses reported here. In order to validate this approach, we probed β values ranging from $1x10^{-3}$ to $1x10^{3}$, but observed no significant improvement in TWIM ATD fit quality (data not shown).

In order to effectively utilize Equation 2-5 to predict TWIM peak widths, estimates of ion velocities are required to provide ion arrival time values (t). A previous description of TWIM ion transport theory[152] provides the following relationship between ion mobility, the structure of the TW electric field, and overall ion transit time:

$$t = \frac{Lv}{K^2 \left(\frac{V}{x}\right)^2} + b \qquad 2-7$$

This expression, upon rearrangement, becomes:

$$t = x^2 \frac{Lv}{K^2 V^2} + b \qquad 2-8$$

Linear regression analysis of Equation 2-8 yields a slope of x^2 and an intercept of b (Figure I-3a and Table I-2). The V/x ratio shown in Equation 2-8 is equivalent to the effective E an ion experiences during its flight during TWIM separation, and can be used in Equation 2-7 to estimate TWIM ion arrival times (Figure I-3b and Table I-3). The x values shown in Equations 2-7 and 2-8 vary as a function of K (Figure I-3c) which necessitates calibration across a range of V and v in order to be able to predict TWIM arrival times.

We have further extended these empirical equations in order to predict TWIM resolution. By dividing Equation 2-5 through on both sides by t^2 we arrive at:

$$\left(\frac{w}{t}\right)^2 = \frac{\gamma + \beta {t_g}^2}{t^2} + \alpha \frac{v}{KE^2} \qquad 2-9$$

IM resolution is typically defined as:

$$R = \frac{t}{w} \qquad 2-10$$

where R is the resolution. Substituting R into Equation 2-9 we produce:

$$R^{-2} = {R_p}^{-2} + {R_d}^{-2} \qquad 2-11$$

where R_p and R_d are the contributions to IM resolution related to the initial ion pulse width and diffusion, respectively. Using Equations 2-9 and 2-11 we can define R_p and R_d individually as:

$$R_p = \frac{t}{\left(\gamma + \beta {t_g}^2\right)^{\frac{1}{2}}} \qquad 2-12$$

$$R_d = \left(\frac{E^2 K}{\alpha v}\right)^{\frac{1}{2}} \qquad 2-13$$

Further, we can use Equations 2-5, 2-7, and 2-10 to obtain a complete R expression:

$$R = \frac{Lv}{K^2E^2\left(\gamma + \beta {t_g}^2 + \frac{\alpha L^2 v^3}{K^5 E^5}\right)^{\frac{1}{2}}} \qquad 2-14$$

Additionally, by setting $\frac{dR}{d\left(\frac{E^2}{v}\right)} = 0$ we find the optimal E^2/v to be:

$$\left(\frac{E^2}{v}\right)_{optimal} = \left[\frac{\alpha L^2}{2K^5\left(\gamma + \beta {t_g}^2\right)}\right]^{\frac{1}{3}} \qquad 2-15$$

In this work, we test the above empirical sets of equations in order to model the TWIM ATD width and resolution for a range of biomolecular ions. In addition, we similarly evaluate the conformational broadening parameter δ for specific peptides and protein complexes.

2.4 Results and Discussion

To test our set of semi-empirical equations describing TWIM peak width, we used a series of Ac-Ala$_n$-Lys peptides , which have previously been observed to adopt rigid helical structures in the gas phase[212,213]. Figure 2-1a and 2-1b shows TWIM-MS data collected for Ac-Ala$_n$-Lys peptides contained within our n=19 sample. Within this sample, we observe a distribution of Ac-Ala$_n$-Lys peptides, with n ranging from 6 to 19. The appearance of n < 19 peptide signals is likely due to the hydrolysis of n=19 peptides under the acidic conditions used to dissolve the original hydrophobic peptide sample, and such conditions mirror those used previously to analyze the gas-phase structures of these sequences[196,212,213]. We observe both $[M+H]^+$ and $[M+2H]^{2+}$ peptide ions in our IM-MS data, with the latter group detected having relatively low signal intensities. Due to the significantly larger signal intensity for the singly-charged peptide ions in our dataset, our detailed peak width analysis focuses on these signals exclusively.

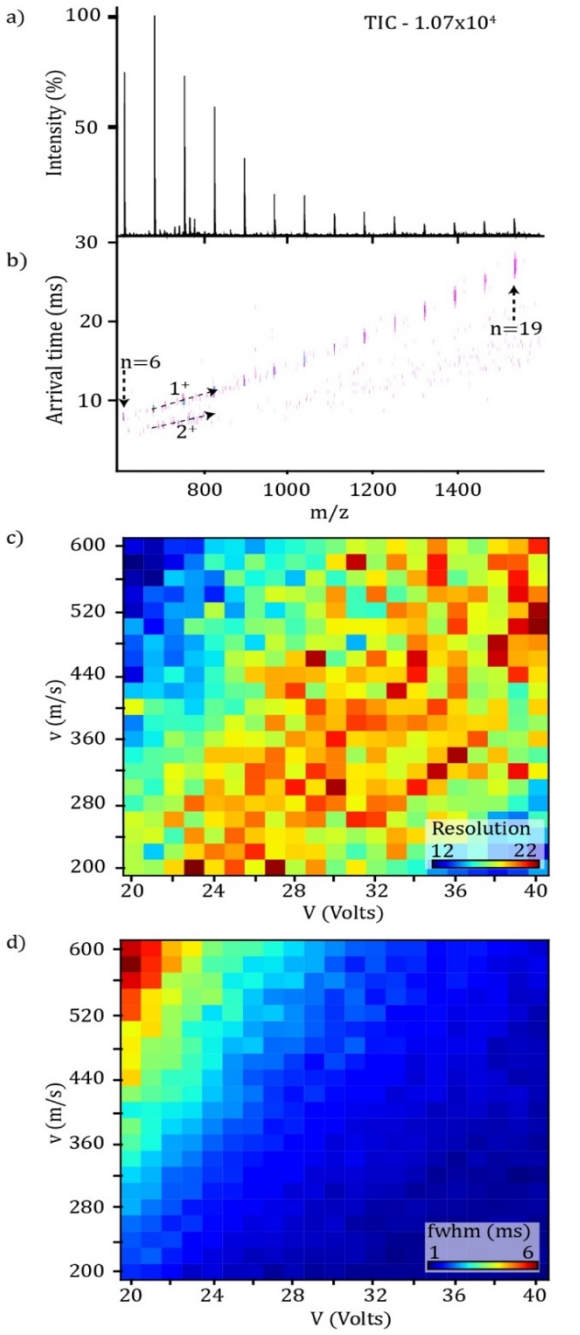


Figure 2-1. IM-MS spectra recorded for a Ac-Ala$_n$-Lys peptide and contour plots of resolution and fwhm as a function of V and v for singly-charged Ac-Ala$_{18}$-Lys peptide. **a)** A mass spectrum recorded for a Ac-Ala$_n$-Lys peptide where n = 19. **b)** A plot of TWIM drift time versus m/z, where ion intensity denoted by color-coded contour scale as indicated, for the same data shown in a. In both **a)** and **b)** we observe both a singly and a doubly charged series for n = 6 – 19 of peptides, with the latter series is detected at substantially lower signal intensities when compared to the former. Contour plots that show the influence of **c)** resolution and **d) fwhm** as a function of **V** and **v** for singly-charged Ac-Ala$_{18}$-Lys peptides, with values for both figures of merit indicated by the color axis shown.

Our peptide TWIM ATD data were acquired for a range of V and v settings at 3.5 mbar of pressure in the TWIM cell. ATDs were extracted for individual peptide systems and fitted with a Gaussian function to obtain a centroid arrival time, full width half max (***fwhm***), and resolution values, which were calculated using Equation 2-10. A contour plot of resolution and ***fwhm*** as a function of V and v is shown in figure 2-1c and figure 2-1d, respectively, for Ac-Ala_{18}-Lys^{1+}. In this data, the ***fwhm*** decreases as V increases and v decreases, reaching a maximum at the smallest V and highest v values probed here. On the other hand, the TWIM resolution trend seen in figure 1c does not mirror the ***fwhm*** trend, and instead reaches optimally large values at a V to v ratio of 0.075-0.125 (Figure I-4), as observed previously for TWIM analyzers[211]. The resolution values observed in this study differ from the ones reported by Giles et al[214], which computes resolution in CCS space. We computed resolution[215] in time space in order to directly link the experimental measurement to the semi-empirical model we devised. After adjusting the effective E field to yield an accurate estimation of centroid arrival times (Equation 2-8, Figure I-3, and Table I-3), we implemented Equations 2-5 and 2-14 to fit our experimental dataset in terms of ***fwhm*** and R, respectively. Figure 2-2a shows a plot of $fwhm^2$ vs vt^2/KE^2 for Ac-Ala_{18}-Lys^{1+} data. Through linear regression, and the application of Equation 2-5, we extract an α value for these data, an umbrella term that includes conformationally-derived peak broadening, with a strong correlation coefficient ($R^2 = 0.9751$) indicating that our model fits the experimental data well. In addition, we utilize the standard error resulting from our linear regression to estimate the error associated with δ value determined in this way. We compared measured R, R_d, and R_p values with computed resolution R values as a function of E^2/v in Figure 2-2b. As E^2/v increases, R_d increases whereas R_p decreases,

as expected. ***R***, which is the weighted sum of R_p and R_d from Equation 2-14 models the functional form observed experimentally. The red dashed line indicates the $E^2/v_{optimal}$ value (from Equation 2-15) where maximum ***R*** is observed is predicted to be observed. A summary of the fitting parameters and $E^2/v_{optimal}$ for all the peptides is shown in Table I-4. Figure 2-2c plots measured and computed ***fwhm*** as a function of E^2/v. Computed ***fwhm*** values agree well with those measured by TWIM, as indicated by the linear regression analysis shown. Specifically, our semi-empirical TWIM width relationship models the ***fwhm*** values for the peptides studied here with an average relative standard deviation of about 5% (Table I-5 and Figure I-5).

In order to quantitatively evaluate the peak broadening observed in our TWIM peptide data related to conformational polydispersity, we evaluated the conformational broadening parameter, δ, by separating this value from the remainder of the α term found in Equation 2-6. When the δ term that describes a TWIM peak approaches a value of 1, the width of that peak is diffusion limited. Any increase in the δ

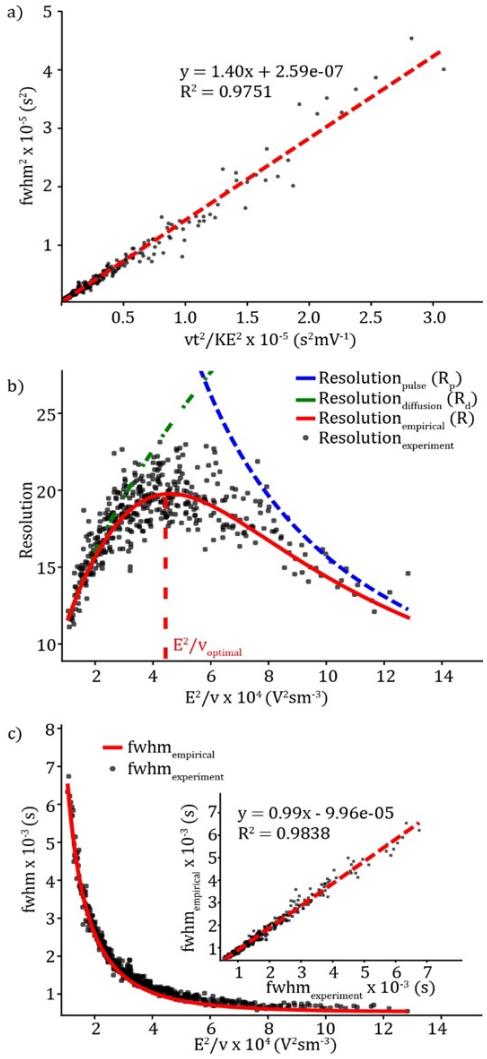


Figure 2-2. Validation of TWIM semi-empirical relationship for resolution and fwhm. **a)** A plot of $\mathbf{fwhm^2}$ as a function of $\mathbf{vt^2/KE^2}$ where linear regression analysis yields $\boldsymbol{\alpha}$ as the slope and $(\boldsymbol{\gamma} + \boldsymbol{\beta}\, t_g^2)$ as the y-intercept, from Equation 5. The correlation coefficient and best fit equation from the linear regression analysis is shown on the plot. **b)** A plot of TWIM resolution as a function of $\mathbf{E^2/v}$. Color coded trends are shown to represent predictions derived from our empirical relationships for diffusion-limited (green) and initial pulse width limited (blue) TWIM resolution. The predicted trend from the complete empirical resolution model is also shown (red solid line), and exhibits a strong correlation to experimentally measured TWIM peak widths (black points). The value for $\mathbf{E^2/v_{optimal}}$ was determined using Equation 16, and is indicated on the plot (red, dashed line). **c)** A plot of **fwhm** as a function of $\mathbf{E^2/v}$, where TWIM data points (in black) are compared against the trend predicted from Equation 3 (red, solid line). The inset shows the correlation between TWIM fwhm values from both our empirical model and experiment, exhibiting an excellent level of correlation and a slope of 0.99.

value needed to describe the width of a TWIM ATD is, therefore, evidence of non-diffusional broadening, specifically those related to the ensemble of gas-phase structures occupied by the ion. Figure 2-3 displays the $\boldsymbol{\delta}$ values for Ac-Ala$_n$-Lys peptides required to fit the TWIM peak widths recorded in our experiments.

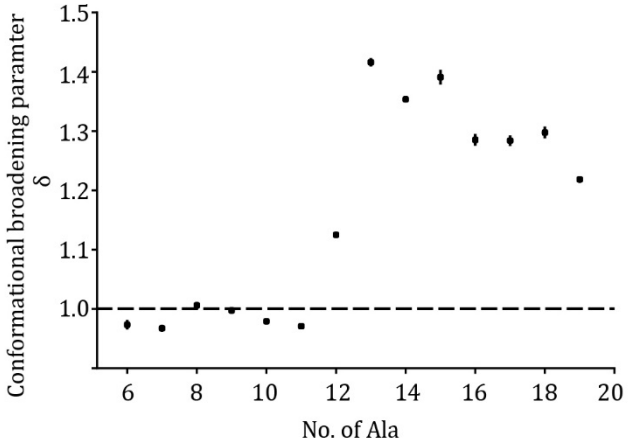


Figure 2-3. Conformational broadening parameters of Ac-Ala$_n$-Lys peptides. A plot of the conformational broadening parameter (**δ**) extracted from our TWIM peak fits as a function of number of alanine residues in the peptide ions measured. The error bars shown represent the experimental error, propagated from the slope of linear regression shown in Figure 2a. The dotted horizontal line at **δ** = 1 indicates the expected d value for diffusion limited TWIM peak widths. Larger **δ** values indicate significant conformational broadening.

We observe $\boldsymbol{\delta}$ values near to 1 for Ac-Ala$_n$-Lys peptides where n = 6 – 11, indicating negligible contributions to observed TWIM peak widths from the peptide conformational ensemble. Furthermore, we observe a sharp transition in the magnitude of the $\boldsymbol{\delta}$ values required to fit our TWIM peak widths at a peptide length of n=12, after which $\boldsymbol{\delta}$ values remain above 1.2 for peptides with lengths of n = 13 – 19. Some of our TWIM peak widths require $\boldsymbol{\delta}$ values slightly below 1 in order to generate accurate fits, likely due to minor inaccuracies in our empirical relationships (Table I-5 and Figure I-5). We observe a large shift in $\boldsymbol{\delta}$ value as the length of the Ac-Ala$_n$-Lys^{1+} peptides increases, indicating significant increases in peptide structural heterogeneity for sequences with n >11. This result correlates well with previous DTIM measurements, where singly-charged Ac-Ala$_n$-Lys ions were generally observed to be helical and rigid. In addition, DTIM widths for similar, singly-charged polyalanine peptides were reported to scale from diffusion-limited values for short sequences, to values 1.5 times higher for n = 20 peptides[196].

In order to model the ensemble of peptide structures present in our TWIM-MS experiments, we utilized MD simulations to generate a large population of Ac-Ala$_n$-Lys peptide structures, equilibrated at 300 K, 400 K, and 500 K. At each temperature, we pooled all low energy structures generated, and plotted these as a histogram to create CCS distributions, which were subsequently fit to Gaussian functions for comparison with TWIM datasets (Figure I-6). Weighted mean values for each distribution were plotted against the number of alanine residues in the peptide analyzed, as shown in Figure 2-4a.

To further evaluate the quantitative agreement between our computed structural ensembles and those observed experimentally by TWIM, we evaluated the weighted mean CCS values generated from both IMPACT and IMOS, and observed good agreement with experimental CCS values, regardless of the temperature used in our MD simulations (Figure I-7). Importantly, we find

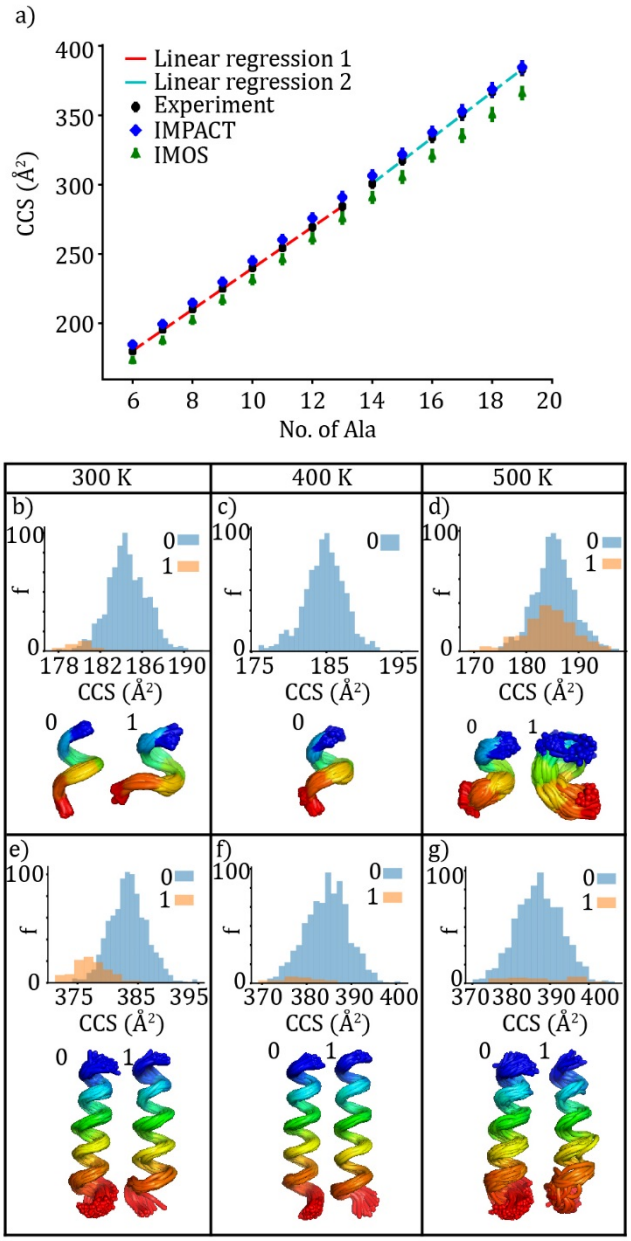


Figure 2-4. Theoretical models of Ac-Alan-Lys peptides. **a)** A trend line analysis of Ac-Ala-Lys CCS data plotted as a function of Ala residues contained within the peptide (black dot points). Theoretical CCS values representing the mean of our MD-generated structural ensembles are also shown, computed using the IMPACT (blue diamonds) and IMOS (green triangles) trajectory methods. Two linear regression models were required to fit the experimental CCS data as indicated by the red and blue dashed lines. **b)**, **c)** and **d)** show the clustered CCS distributions for Ac-Ala$_6$-Lys peptide ions at 300 K, 400 K, and 500 K, respectively. **e)**, **f)**, and **g)** show similarly clustered CCS distributions for Ac-Ala$_{19}$-Lys peptide ions at 300 K, 400 K, and 500 K, respectively. Clusters are numbered as 0 and 1, as referred to in the text. Ensembles of overlaid structures extracted from clusters are also shown below each CCS distribution plot, and are labelled with respect to their cluster of origin.

that two linear regression models provide the best fit for the experimental data shown in Figure 2-4a (Figure I-8) with a first linear trend covering peptides with n = 6-13and a second trend covering n = 14-19. Slope values for these two trend lines are 14.9 $Å^2$/Alanine for shorter peptides and 16.5 $Å^2$/Alanine for those that are longer. We observe similar variance in the fit when the two trends described above are restructured so that the first linear regression covers peptides with n=6-12, and the second covers those with n = 13-19 (Figure I-8), clearly indicating a shift in CCS increase per-alanine value observed at either n = 12 or 13. Importantly, the transition point recorded for our CCS trend lines occurs at the same peptide lengths where we observe a sharp increase in $\boldsymbol{\delta}$ values in Figure 2-3.

In pursuit of a deeper structural interpretation for our TWIM data, we interrogated our MD results by classifying the resulting structural ensembles in terms of the conformational families produced during our simulations. Hierarchical clustering on a reduced dataset of ~1000 structures extracted from each constant temperature run was used to classify structural families. Families accounting for greater than 2% of the total population were considered for assessing dynamics (Tables I-7 and I-8). Figures 2-4b, 2-4c, and 2-4d show our classification results for n = 6 peptide models at 300K, 400K, and 500K, respectively, whereas Figures 2-4e, 2-4f, and 2-4g show classifications for n = 19 peptides similarly structured with respect to temperature. In general, our simulations reveal broader ranges of ion CCS values for longer peptides, in agreement with our TWIM peak width data. Our cluster analysis indicates that while primarily helical, structural heterogeneity occurs in Ac-Ala_n-Lys peptide ions at their C-termini (Figure I-9 and Tables I-6, I-7, and I-8). Notably, MD simulations for shorter peptide sequences, from n = 6-10, contain a structural family where helicity is not maintained throughout the peptide. For example, cluster 1 in 500K dataset for singly-charged Ac-Ala_6-Lys exhibits a more random coil type conformation

(Figures 2-4d and I-9). Despite these structural differences, random coil and helical peptide families have very similar mean CCS values (Tables I-6, I-7, and I-8). For larger peptides, the helical structural families in our MD simulations appear to more completely dominate the structural landscape, with significant variations in the C-termini (Figure 2-4e, 2-4f, 2-4g, and Figure I-9). While the widths of our MD-generated ensembles in CCS space to not quantitatively correlate with our experimental peak widths from TWIM data, such agreement is not likely given the unknown temperatures of the ions measured and the relative simplicity of our simulations (Figure I-10)[152,216,217]. Overall, however, our MD simulations point to the conformational diversity of helical peptide ion C-termini as the main driver of the increased $\boldsymbol{\delta}$ values observed in Figure 2-3.

The ultimate aim of our TWIM width analysis workflow is the rapid assessment of protein structural heterogeneity. To demonstrate the capabilities of our TWIM method, we measured $\boldsymbol{\delta}$ values for Avidin, a 64 kDa homo-tetrameric protein complex, as a function of the concentration of added CXL reagent in solution. We anticipate that CXL reactions will, in general act to constrain the Avidin structure, yielding a population of gas-phase ion structures of reduced structural heterogeneity, and thus decreased $\boldsymbol{\delta}$ values. Figure 2-5 shows $\boldsymbol{\delta}$ values recorded

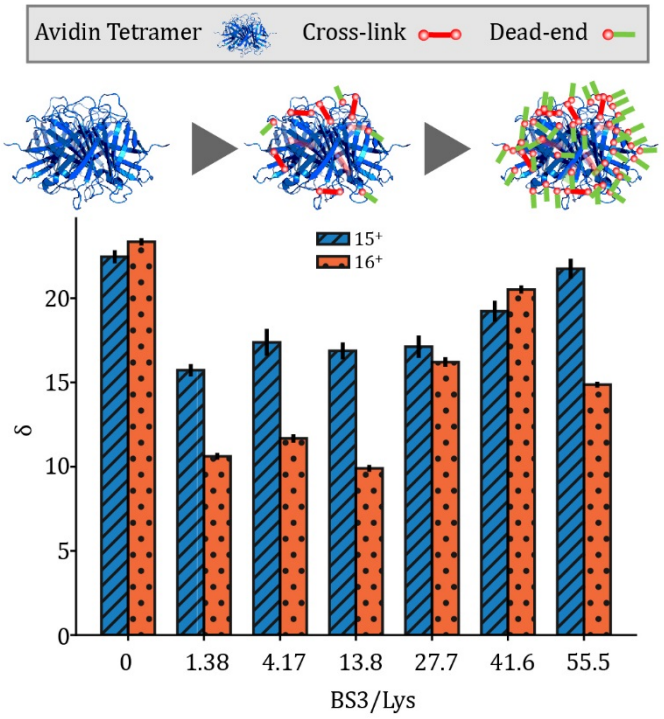


Figure 2-5. Changes in conformational broadening parameter as a function of cross-links. Bar graph of **δ** values needed to fit TWIM peak widths recorded for unmodified and BS3 cross-linked Avidin tetramer 15+ (blue) and 16+ (red) ions. We interpret the initial decreases in **δ** values upon treatment with CXL agents as evidence of rigidified protein structures, whereas increases in **δ** values observed at larger BS3 excesses are likely due to the prevalence of dead-end cross-links under

for Avidin as a function the ratio between BS3, a CXL reagent that targets primary amines, and the available Lys residues within the Avidin sequence. Remarkably, we note $\boldsymbol{\delta}$ values for Avidin ions between 10-20 times those observed for Ac-Ala_n-Lys peptide ions, indicating a dramatically increased level of structural polydispersity for the protein complex. As the BS3/Lys ratio is increased, we observe a significant decrease in the $\boldsymbol{\delta}$ values required to accurately fit our TWIM peak widths, indicative of a restrained population of Avidin ion structures across both charge states probed. We note that different $\boldsymbol{\delta}$ values are required to fit different Avidin charges states, and that changes in Avidin TWIM peak width appear non-uniform across those Avidin charge states probed here, indicating that each protein complex charge state is comprised of a unique structural ensemble. We observe minimal $\boldsymbol{\delta}$ values BS3/Lys ratios of 1.38 and 13.8 for 15+ and 16+ ions respectively. As the BS3/Lys ratio is increased beyond 13.8, we further observe an increase in $\boldsymbol{\delta}$ values for both charge states, indicating an increase in structural polydispersity, likely driven by the prevalence of dead-end reaction products known to dominate under such reaction conditions previously[218–220]. Overall, our Avidin results illustrate the performance of our empirical TWIM peak width analysis approach. The changes in $\boldsymbol{\delta}$ we observe match our expectations for CXL modified Avidin complex ions, as well as revealing new quantitative insights into gas-phase protein complex structures.

2.5 Conclusions

Dynamic ensembles, rather than static structures, dictate the function of active biomolecules. The TWIM width analysis approach presented here provides a rapid assay of protein structure heterogeneity. We develop and validate a semi-empirical relationship that accurately models TWIM ATD widths and resolutions across a wide range of instrument settings. In addition, we use this workflow to examine the structure of peptide and protein complex model systems. By

extracting a conformational broadening parameter, $\boldsymbol{\delta}$, from our empirical formulae, we quantified the conformational broadening for a series of Ac-Ala_n-Lys peptides, with n = 6 – 19. Our width analyses allowed us to classify these peptide ions into two families in terms of their structural heterogeneity, with n=6-11 having $\boldsymbol{\delta}$ values indicating a diffusion-limited peak width, and $n \geq 12$ exhibiting $\boldsymbol{\delta}$ values congruent with more significant conformational broadening. Furthermore, we utilized MD simulations in combination with hierarchical clustering analysis in order to locate the source of the structural polydispersity in our TWIM peptide data to the C-terminal regions of the longer sequences studied. Finally, we implemented our TWIM width analysis approach to study the influence of chemical cross-linking on the structural ensemble occupied by the Avidin tetramer, recovering large $\boldsymbol{\delta}$ values that decrease sharply upon CXL treatments, followed by increases at when CXL agent excesses drive the formation of dead-end reaction products. Our quantitative TWIM peak width analyses also detected charge state dependent effects, indicating the presence of distinct structural ensembles for Avidin tetramers previously hidden within iso-CCS TWIM features.

We envision that our TWIM peak width analysis approach will enable a broad spectrum of applications in protein structural biology and biophysics. Examples include, but are not necessarily limited to, the study protein aggregation, evaluating the functional ensembles of intrinsically disordered proteins (IDPs), and rapidly monitoring the effect of ligand binding on proteins targets. Generating accurate estimates of biomolecular ensembles remains a challenge for computational biochemistry, and such difficulties extend directly to our ability to accurately model structural distributions *in vacuo* in a manner that corresponds to our TWIM-MS experiments. Furthermore, untangling the influence of charge state on the biomolecular ensembles quantified in our Avidin measurements will likely require the long-term application of charge manipulation, high resolution

IM, and tandem IM experiments in combination with sophisticated MD simulations. However, it is clear from the data presented here that our empirical relationships describing TWIM peak width performs well across a wide range of protein analytes, and is accessible rapidly (within seconds) of sample introduction. Future applications of TWIM, in combination with other gas-phase structural biology approaches, will undoubtedly lead to an improved understanding of the structural ensembles associated with function biomolecules.

2.6 Acknowledgements

We thank Kevin Giles and David Langridge at Waters for providing initial SIMION files to build G2 model and extract voltage and electric field. We further thank Kevin Giles for helpful discussions with TWIM theory. We would also like to thank Efrosini Artikis from the Brooks group at University of Michigan for helpful MD discussions. Finally, we gratefully acknowledge funding from the National Science Foundation Division of Chemistry under Grants 1808541 and 1253384 (with co-funding from the Division of Molecular and Cellular Biosciences) for supporting our efforts in TWIM theory development.

Chapter 3 A novel ion pseudo-trapping phenomenon within traveling wave ion guides

(Equal contributions by: Sugyan M. Dixit, David Langridge (Waters Corporation), and Keith Richardson (Waters Corporation))

3.1 Introduction

Ion mobility (IM) separation has been used for the physical and structural characterization of molecules for nearly 50 years[221–225]. The utility of IM is maximized when coupled to mass spectrometry (MS), as both ion size and m/z can be recorded rapidly in a manner that has proven to be enabling for areas ranging from food analyses to tissue imaging.[226–228] In recent years, IM-MS has gained traction as an approach for the system-wide analyses of cellular constituents[229–232] and in structural biology[124,189,233] due to its ability to obtain orientationally averaged collision cross-section (CCS) values used as restraints for bimolecular modeling[145,193,234]. Moreover, the development of commercial instruments, including those that incorporate drift tube IM (DTIM), differential mobility, traveling wave IM (TWIM), trapped IM spectrometry (TIMS), and structures for lossless ion manipulation (SLIM) devices, has spurred the widespread use of IM-MS over the past decade[125].

The introduction of TWIM technology[197] coupled to MS in a quadrupole/TWIM/oa-TOF platform [148] acted to significantly expand the user base of IM-MS technology[190]. A TWIM device is comprised of a radio frequency (RF) ion guide with repeating DC pulses applied to a pair (or two pairs) of ring electrodes sequentially in the device creating a traveling wave electric field (E). Higher mobility ion species are overtaken by the waves, resulting in a “roll-over” event, less often

than low mobility species. The latter ions, thus, travel more slowly through a TWIM separator than the former, enabling a mobility-based separation[151,197,235–237]. However, due to the dynamic field and complex ion motion involved, obtaining an ion mobility value directly from observed TWIM arrival time distributions (ATDs) remains challenging. This necessitates the calibration of TWIM drift times using reference CCS values obtained from DTIM. Several studies have described the development of TWIM CCS calibration strategies using data acquired for various standards including small drug-like molecules[238], sugars[239], lipids[240], polynucleotides[241], peptides[242], proteins, and protein complexes[243]. Although several pioneering studies have developed theoretical frameworks that have enabled deeper insights into understanding the ion motion in TWIM[152,153,198,199], there are still deficiencies that make it challenging to predict characteristics of TWIM ATDs as well as obtain CCSs directly from TWIM measurements.

Here we observe and theoretically describe a novel pseudo-trapping effect within TW ion guides with implications for both TWIM separations and CCS determinations. The effect of pseudo-trapping can be observed experimentally as a sharp increase in IM resolving power, defined as $\mathbf{t}/\mathbf{\Delta t}$, where $\mathbf{t}$ and $\mathbf{\Delta t}$ are the centroid and fwhm of TWIM ATDs, respectively. Ion trajectory simulations during pseudo-trapping show a specific and repetitive pattern of motion causing ions that have different mobilities to travel with the same mean velocity. We present our theoretical frameworks to predict and describe such pseudo-trapping in TWIM. Theory and simulations suggest that TW waveforms that are not perfectly sinusoidal and smoothly moving can exhibit the pseudo-trapping behavior we describe. We conclude by describing the impact of pseudo-trapping on the CCS calibration of TWIM measurements.

3.2 Methods

3.2.1 Samples

Ubiquitin (U6253), cytochrome c (C2506), β-lactoglobulin (L7880), and D, L polyalanine (P9003) were purchased from Sigma-Aldrich, USA. Protein samples were prepared in 200 mM ammonium acetate solution at a concentration of 5 μM. Polyalanine was dissolved in 49.5%/49.5%/1% water/acetonitrile/acetic acid solution at 1 mg/ml final concentration.

3.2.2 TWIM-MS

All data was collected using a Synapt G2 instrument platform (Waters, Millford MA) [214,244]. Briefly, ions were generated using nano electrospray ionization (nESI) and then pulsed into the TWIM cell. TWIM device was operated at a pressure of 3.5 mbar with a combination of both He and N_2 gas present. TWIM-MS data was acquired at values of wave amplitude (**V**) ranging from 20 to 40 V and values of wave velocity (**v**) ranging from 200 to 1000 m/s. After ions transit the TWIM cell, they are transferred to an orthogonal acceleration (oa) time of flight (ToF) mass analyzer through a TW ion guide. TWIM ATDs are recorded through synchronization of oa-ToF acquisition with the gated release of ions from the trap into the TWIM cell. CCS was calibrated at all **V** and **v** conditions using methods previously described[155].

3.2.3 SIMION Simulations

SIMION 8.1[200] was used to model ion transport within the TWIM cell[197]. By fixing ions to only axial direction, and removing RF confinement and velocity relaxation, a simplified 1 dimensional (1D) model was employed to study the ion traces and investigate the effects of anharmonicity and wave stepping. A full off-axis 3 dimensional model (3D) with RF confinement,

SDS gas model for diffusion, and velocity relaxation was also used. About ~1000 or more trajectories were created to accumulate a sufficient distribution of drift time values for analysis.

3.2.4 Data Analysis

TWIM ATD data was extracted using TWIMExtract[206]. ATDs were fitted with a Gaussian function to obtain **t** and **Δt**. CCS was calibrated using a power-law function as described previously[155]. Data were analyzed using python, numpy, and scipy[204,205] with in-house written scripts. Matplotlib[207] was used to generate all the plots.

3.3 Theory: A predictive model for pseudo-trapping in TWIM

For a smooth sinusoidal[237] wave with wavelength **λ**, roll-over period **T** can be described as

$$\mathbf{T} = \frac{\boldsymbol{\lambda}}{\mathbf{v}} \frac{\mathbf{1}}{\sqrt{\mathbf{1} - \boldsymbol{\gamma}^{\mathbf{2}}}} \qquad 3-1$$

where **v** is the wave velocity,

$$\boldsymbol{\gamma} = \frac{\mathbf{kKV_0}}{\mathbf{v}} \qquad 3-2$$

k is the wave number, **K** is ion mobility, and $\mathbf{V_0}$ is the amplitude of the wave potential.

Pseudo-trapping appears to occur when **T** is an integer number **m** of wave steps

$$\mathbf{T} = \mathbf{m}\mathbf{t_s} \qquad 3-3$$

where

$$\mathbf{t_s} = \frac{\mathbf{d_s}}{\mathbf{v}} \qquad 3-4$$

$\mathbf{d_s}$ is the step distance.

Rearranging Eq. 3-3 and substituting **T** and $\mathbf{t_s}$ from Eqs. 3-1 and 3-4, respectively, we get

$$\mathbf{m} = \frac{\mathbf{T}}{\mathbf{t_s}} = \frac{\boldsymbol{\lambda}}{\mathbf{d_s}} \frac{\mathbf{1}}{\sqrt{\mathbf{1} - \boldsymbol{\gamma}^{\mathbf{2}}}} = \frac{\mathbf{n}}{\sqrt{\mathbf{1} - \boldsymbol{\gamma}^{\mathbf{2}}}} \qquad 3-5$$

where

$$\mathbf{n} = \frac{\boldsymbol{\lambda}}{\mathbf{d_s}} \qquad 3-6$$

Therefore, we predict pseudo-trapping in TW ion guides at

$$\boldsymbol{\gamma} = \sqrt{\mathbf{1} - \frac{\mathbf{n^2}}{\mathbf{m^2}}} \ \mathbf{only\ for\ m > n} \qquad 3-7$$

3.4 Theory: A model system describing pseudo-trapping in TWIM

In the absence of velocity relaxation, the following differential equation describes the ion's motion in a general time-dependent electric field in TW reference frame [237]

$$\frac{\mathbf{dz}}{\mathbf{dt}} = \boldsymbol{\gamma}\mathbf{sinz} - \mathbf{1} \qquad 3-8$$

where

$$\mathbf{z} = \mathbf{kx} - \boldsymbol{\omega}\mathbf{t} + \ \boldsymbol{\varphi} \qquad 3-9$$

k is the wavenumber, **ω** is the angular frequency, and **φ** is a phase constant.

In this reference frame, wave steps look like a time dependent phase shift as following

$$\frac{\mathbf{dz}}{\mathbf{dt}} = \boldsymbol{\gamma}\ \mathbf{sin}\big(\mathbf{z} + \boldsymbol{\delta}(\mathbf{t})\big) - \mathbf{1} \qquad 3-10$$

where **δ(t)**, a sawtooth function, can be expanded as a Fourier series

$$\boldsymbol{\delta}(\mathbf{t}) = \frac{\boldsymbol{\pi}}{\mathbf{4}} - \frac{\mathbf{1}}{\mathbf{2}}\sum_{\mathbf{n=1}}^{\infty}\frac{\mathbf{1}}{\mathbf{n}}\mathbf{sin4nt} \qquad 3-11$$

By taking only the first term from the expansion in Eq. 3-11, we can use **δ(t)** to describe a wave that oscillates back and forth in the TW frame:

$$\boldsymbol{\delta}(\mathbf{t}) = -\boldsymbol{\zeta}\mathbf{cos}\boldsymbol{\Omega}\mathbf{t} \qquad 3-12$$

where **ζ** is the amplitude of oscillation in **z**, and **Ω** is the angular frequency.

By adding anharmonicity to Eq. 3-8 the following describes our model system capable of reproducing pseudo-trapping in TW reference frame:

$$\frac{\mathbf{dz}}{\mathbf{dt}} = \boldsymbol{\gamma}\left(\mathbf{sin}\left(\mathbf{z} + \boldsymbol{\delta}(\mathbf{t})\right) + \mathbf{asin3}\left(\mathbf{z} + \boldsymbol{\delta}(\mathbf{t})\right)\right) - \mathbf{1} \qquad 3-13$$

where **a** is the anharmonicity scaling factor. The introduction of anharmonicity in our model allows us to investigate the role of non-ideal TW form in pseudo-trapping.

3.5 Results and Discussion

During our previous efforts to establish an empirical model for describing TWIM ATDs[245], we observed an aberrant trend in **t/Δt** at certain TW amplitudes and velocities. Figure 3-1a, b, and c show **t/Δt** as a function of **γ** for ubiquitin 5^+, cytochrome c 6^+, and β-lactoglobulin 8^+ ions, respectively. The trend seen in Figure 3-1 agrees well with our previous model [245] except at **γ** values of 0.67 and 0.82 , where we observe a significant increase in **t/Δt,** which translates to a 63%, 100%, and 200% increase in apparent TWIM resolving power for ubiquitin 5^+, cytochrome c 6^+, and β-lactoglobulin 8^+ ions, respectively, at a **γ** value of 0.67. Enhancements in apparent TWIM resolving power are lessened at a **γ** value of 0.82, with **t/Δt**

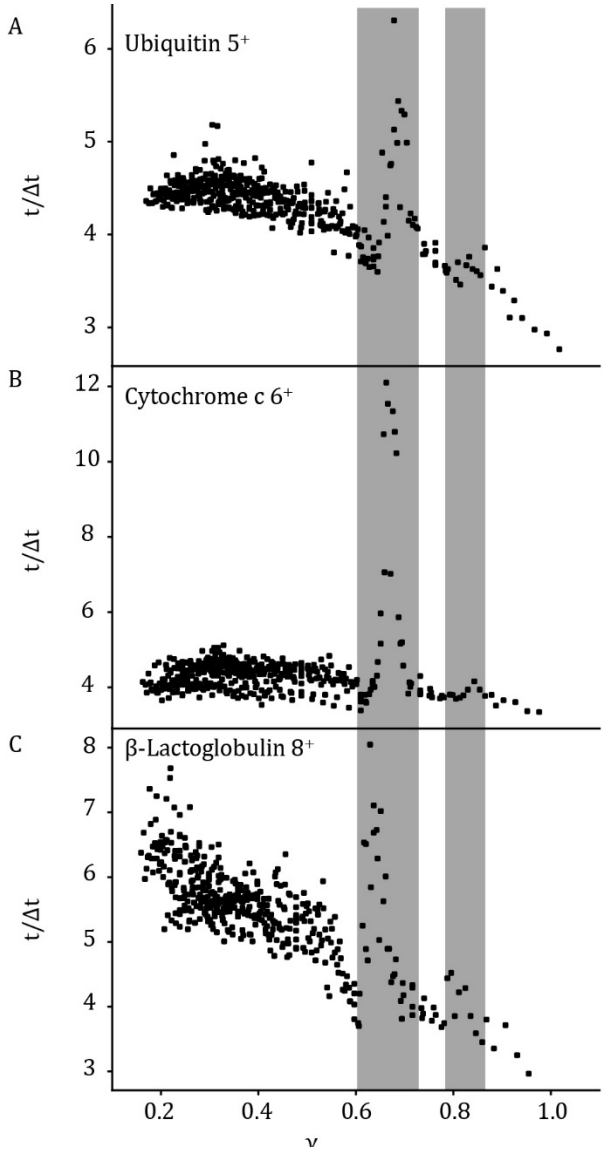


Figure 3-1. Experimental observation of spikes in t/Δt as a function of γ. A plot of t/Δt as a function of γ for A) ubiquitin 5+, B) cytochrome c 6+, and C) β-lactoglobulin 8+ species. The grey shaded regions indicate the spike in t/Δt at γ values of approximately 0.67 and 0.82.

increases of less than 20% in all cases at. This, to our knowledge, is the first observation of such abrupt changes in TWIM resolving power. In order to begin ascribing a mechanism to the observations shown in Figure 3-1, we first invested the role of potential ion heating induced structural changes on TWIM resolving power values at the **γ** values indicated above. We calculated the field strengths necessary to remain within "low-field" conditions for the ions investigated in this study using previously described formulae [153]. The large number of rotation and vibrational degrees of freedom found in the protein ions studied here make it highly unlikely for significant ion heating to occur under our experimental conditions. We therefore interpret in the abrupt increases in **t/Δt** observed in Figure 3-1 as evidence of pseudo-trapping within TW ion guides.

The numerical models and simulations described above are able to successfully reproduce the changes in **t/Δt** displayed in Figure 3-1. Figure 3-2a shows γ values computed using Eq. 3-5 that are predicted to produce apparent increases in TWIM resolving power due to pseudo-trapping. Specifically, when the appropriate number of wave steps per unit wavelength are input into the model (a value of 4 for the Synapt G2 platform used to collect the data shown

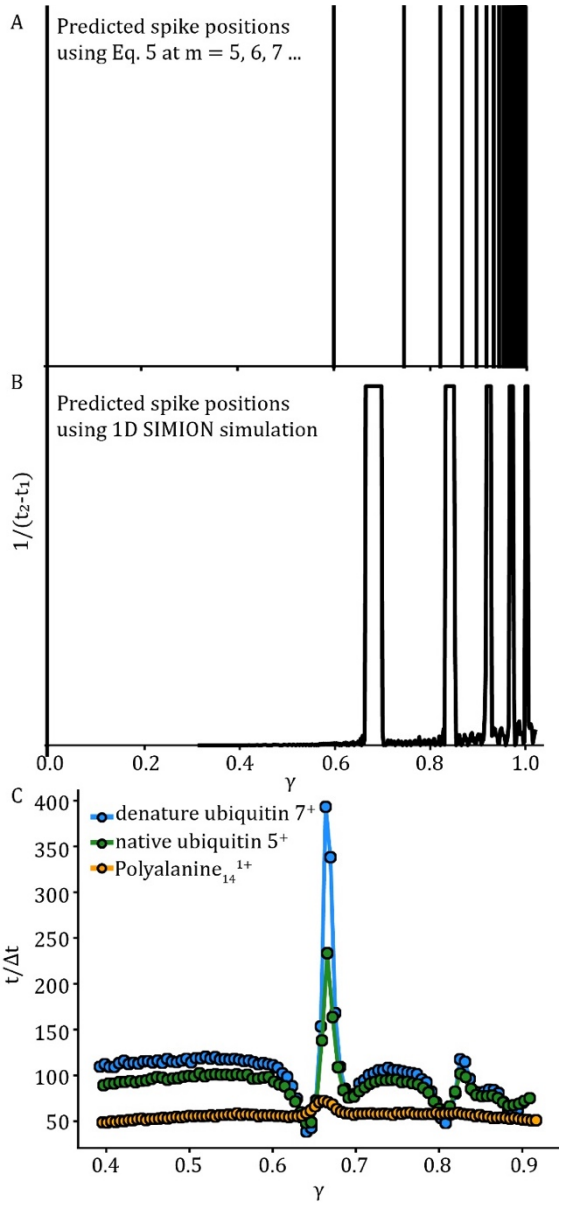


Figure 3-2. Prediction of spike using numerical and simulation models. A plot of spike positions in γ using Eq. 5. (A) and using 1D SIMION simulation (B) with a 3 mm wave step for denatured cytochrome c 19+ ion. C) A plot of t/Δt as a function of γ using 3D SIMION simulation for denatured ubiquitin, native ubiquitin, and poyalanine$_{14}^{1+}$ ions.

in Figure 3-1), we predict that pseudo trapping occurs at $\boldsymbol{\gamma}$ values of 0.60, 0.74, and 0.82, where **m** = 5, 6, and 7, respectively, from Eq. 3-5. To perform similar predictions using our simplified 1D SIMION simulations using a 3 mm step wave, two ions of identical **K** were created 0.5 mm apart at the start of the simulation run, tracked over a range of TW voltages. Figure 3-2b plots the inverse of the time difference between these two test ions ($1(/t_2\text{-}t_1)$) over a range of $\boldsymbol{\gamma}$ values. At $\boldsymbol{\gamma}$ values similar to those discussed above in the context of Figure 3-2a, the difference in drift time between the two test ions becomes small, causing large increases in $1/(t_2\text{-}t_1)$. Predicted $\boldsymbol{\gamma}$ values for ion pseudo-trapping in Figures 3-2a and 3-2b match those observed experimentally (Figure 3-1). Finally, using a 3D velocity relaxed SIMION model, we obtained simulated distributions of ion arrival times for denatured ubiquitin 7^+, native ubiquitin 5^+, and polyalanine$_{14}{}^{1+}$ ions. A plot of $\mathbf{t}/\boldsymbol{\Delta}\mathbf{t}$ values obtained from these simulations as a function of $\boldsymbol{\gamma}$ for is shown in Figure 3-2c. As above, we observe dramatic increases in $\mathbf{t}/\boldsymbol{\Delta}\mathbf{t}$ at $\boldsymbol{\gamma}$ values that match experimental observations from Figure 3-1.

In order to investigate ion pseudo-trapping within TW ion guides in additional detail, we utilized our 1D SIMION model to produce ion trajectories for cytochrome c 19^+ ions at TW amplitudes ranging from 22 V to 27 V. Figure 3-3a shows the ion arrival time as a function of TW amplitude for these 1D simulations. As TW amplitude is increased, we observe a general decrease in ion transit time. However, we observe a plateau in ion residence time between TW amplitude values of 24.3 V to 25.6 V. Over this range of TW amplitudes ions are traveling at the same velocity in our simulations and are, thus, pseudo-trapped within the guide. We examined ion trace data at simulated TW amplitudes positioned either within or outside of this plateau region (Figure 3-3b and 3-3c). We observe that the density of ion positions is greater on the crest of the TW field front than at other positions within the TW reference frame, eluding to the fact that ions are

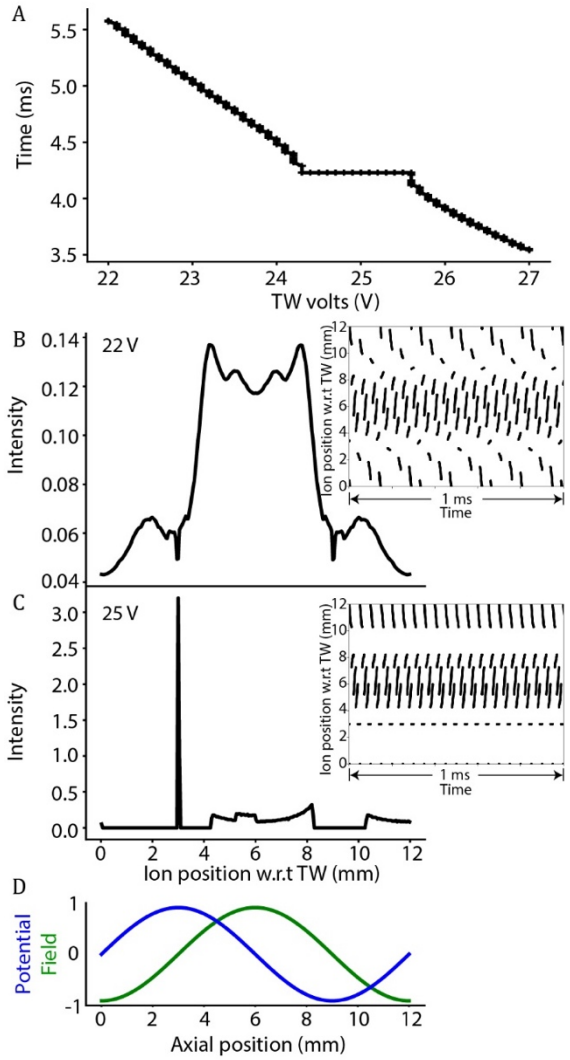


Figure 3-3. Investigation of pseudo ion trapping region using ion trajectory simulations. A) A plot of ion transit time as a function of TW volts. The plateau in time indicates the pseudo-trapping. B) and C) are histogram plots showing ion intensity as a function of ion position w.r.t. TW frame at TW amplitudes of 22 V and 25 V, respectively. The insets in these figures show the ion traces with a plot of ion position w.r.t. TW frame as a function of time. D) A plot of TW field and potential as a function of axial position.

experiencing roll-over events as is expected in TWIM separation (Figure 3-3b). However, during pseudo-trapping, we observe a sharp peak of ion intensity at 3 mm with respect to (w.r.t.) the TW frame. By interrogating the ion trace as a function of time (Figure 3-3c inset), we observe ions experiencing a repeating pattern of ion motion, where 20% of the ion population inhabits specific position in the TW reference frame. It is clear that ion motion captured during our simulations can adequately explain our experimental observation (Figure 3-1), as the repeating pattern of ion motion we identify in Figure 3-3c leads our test ions to both travel with the same mean velocity across a range of TW potentials and diminished diffusional broadening due to a reduction in the net spatial distribution of ion positions within the TW ion guide.

In order to understand the origin of pseudo-trapping in TWIM, we used our simulations and theoretical model to investigate the contribution of the TW field to the overall phenomenon. Figure 3-4a plots the average period **T** (using Eq. 3-13) as a function of **γ** employing various anharmonicity scaling factors (**a**). When a perfectly sinusoidal waveform is used (**a**=0), **T** increases smoothly with increasing **γ** (Figure 3-

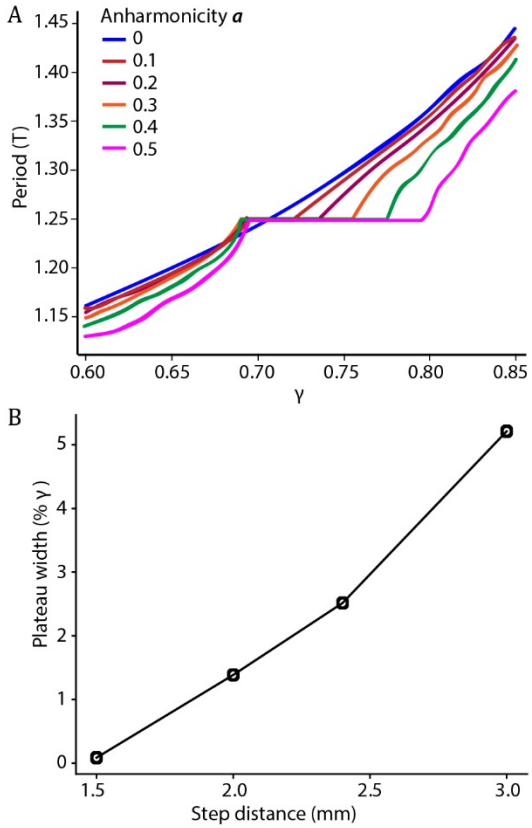


Figure 3-4. Effect of anharmonicity and wave step distance on the average period T. A) A plot of average period T as a function of γ at various anharmonicity scaling factor a values generated using Eq. 3-13. The width of the plateau in γ increases as a function of a. B) Plot showing plateau width in % γ as a function of wave step distance using 1D SIMION simulation.

4a). However, when anharmonicity is increased, a plateau in **T** develops over a range of **γ** values (Figure 3-4a). The range of **γ** values for which **T** remains constant also increases with increasing **a** (Figure 3-4a). From our 1D SIMION simulation, we investigated the plateau in T as a function of TW step distance (Figure 3-4b) and observed that the plateau width increases as a function of increasing step distance (Figure 3-4b). Thus, both our theoretical model and simulations suggest that non-ideal TWs, that lack a perfectly sinusoidal shape or a smoothly moving wave, contributes the pseudo-trapping of ions.

Finally, we investigated how pseudo-trapping of ions in TWIM affects calibrated CCS values. We determined experimental CCS values of cytochrome c 6^+ and β-lactoglobulin 8^+ ions using calibration over a broad range of **γ** values. Figure 3-5a shows the correlation coefficient R^2 from the CCS calibration as a function of **γ**. The R^2 values determined are in excess of 0.99 at **γ** values either insufficient or greater than those to drive TW pseudo-trapping, but R^2 decreases to 0.956 at a **γ** value of ~0.67, which is associated with pseudo-trapping (Figure 3-5a). This indicates that the fitness of TWIM CCS calibration is adversely affected when calibrant ions undergo pseudo-trapping. Figure 3-5b plots the error in predicted CCS values for cytochrome c 6^+ and β-lactoglobulin 8^+ ions as a function of **γ** values. The experimental CCS values are within 1.5% of reference values for cytochrome c 6^+ and β-

lactoglobulin 8^+ ions under all conditions not associated with peudo-trapping (Figure 3-5b). However, during pseudo-trapping, the magnitude of CCS deviation increases up to 3% and 5% for cytochrome c 6^+ and β-lactoglobulin 8^+ ions, respectively (Figure 3-5b). We also performed CCS calibrations using ion transit times extracted from our 3D SIMION model. The trends in CCS deviation from these simulated data matches well with that observed for the experimental data as shown in Figure 3-5b. This indicates that pseudo-trapping can significantly alter ion transit times in TWIM leading to CCS calibrations of reduced accuracy.

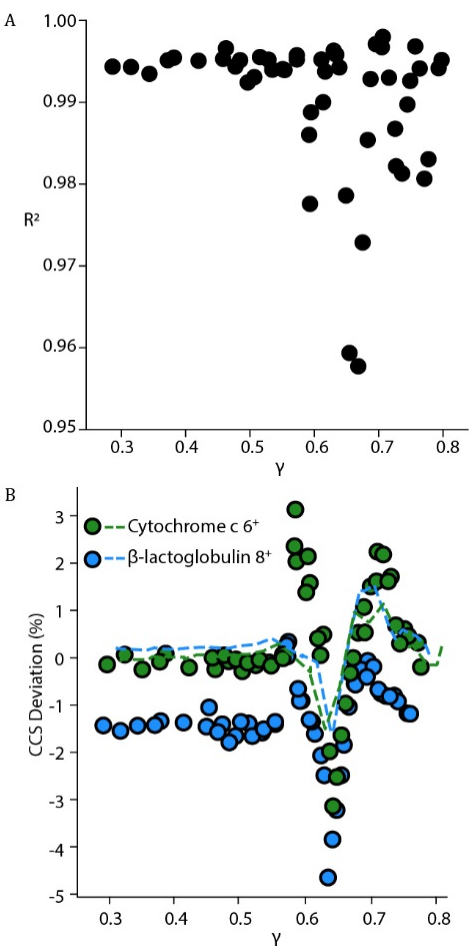


Figure 3-5. Effect of pseudo ion trapping in TWIM CCS calibration. A) A plot of R2 vs. γ showing the goodness of fit for calibration across a range of TW conditions. R^2 decreases during pseudo-trapping in TWIM. B) A plot of deviation (%) in CCS for predicted values from calibration vs. **γ** for cytochrome c 6^+ (green) and β-lactoglobulin 8^+ (blue) species where dots are experimental data and dashed lines are simulated data. The magnitude of CCS deviation increases during pseudo-trapping.

Overall, this work suggests that TW pseudo-trapping of ions should be avoided in order to obtain reliable and accurate ATDs from TWIM analyzers. Under fixed TW amplitude and velocity conditions as long as $\mathbf{v/V_{req} > 370\ K}$, where $\mathbf{V_{req}}$ is the desired wave amplitude, is satisfied, pseudo-trapping effects are not observed. Pseudo-trapping is also eliminated when TWIM is operated under ramped TW amplitude or velocity conditions.[246]

3.6 Conclusions

In this study, we present a new development in TWIM theory that identifies conditions where pseudo-trapping occurs within TW ion guides. We observe a compression of TWIM ATD widths causing a spike in $\mathbf{t/\Delta t}$ values under field conditions predicted to support pseudo-trapping. Our numerical models and simulations can reproduce experimental results where we observe dramatic increases in apparent TWIM resolving power. By analyzing ion traces from simulations, we observe that ions follow a repetitive pattern of motion during pseudo-trapping, spending one out of every five wave steps at a position of 3 mm w.r.t. TW frame during pseudo-trapping. This leads to ions having slightly different mobilities, or those experiencing slightly different TW potentials, to travel with the same mean velocity. We find that non-ideal waveforms, i.e. not perfectly sinusoidal and smoothly moving, exhibit pseudo-trapping at high values of $\boldsymbol{\gamma}$. During pseudo-trapping, the centroid transit time, as well as the width of ion ATDs, are altered, leading to a reduction in the accuracy of calibrated CCS values from TWIM analyzers. Thus, in order to use TWIM devices to accurately assess CCS, pseudo-trapping conditions must be avoided.

By combining this work with our previous efforts to model TWIM ATDs, we are now approaching a nearly complete understanding of the TWIM peak widths. In the future, we envision developing workflows for improved CCS calibration as well as the direct measurement of $\mathbf{K}$ from the TWIM platforms. Such capabilities would leverage a growing population of TWIM analyzers[150,247] to perform ever more accurate assessments of ion CCS for applications in complex mixture analysis and structural biology.

Chapter 4 An improved calibration approach for traveling wave ion mobility spectrometry: robust, high-precision collision cross sections

(Equal Contributions by: Sugyan Dixit, David Langridge (Waters Corporation), and Keith Richardson (Waters Corporation))

4.1 Introduction

Ion mobility (IM), a gas phase technique that separates analyte ions according to their size and charge, is a central technology in modern analytical chemistry, enabling a wide array of both quantitative and qualitative measurements in areas including, but not limited to, atmospheric chemistry, threat detection, drug discovery, food science, proteomics, and structural biology. In many of these areas, IM is coupled directly to mass spectrometry (IM-MS)[248–251], creating a two-dimensional separation technology with synergies that drive improved analyses of complex mixtures derived from a variety of sources. In time-dispersive IM-MS instrument platforms, ions are pulsed into the IM cell containing neutral gas at a controlled pressure. Under the influence of a weak electric field (E), ions undergo several collisions with gas molecules causing them to separate according to their ion-neutral collision cross-sections (CCS).[127] The time required for an ion to traverse such an IM separator (t_d) can be used to calculate its CCS. Many IM platforms are capable of providing CCS values, including: drift tube IM (DTIM), travelling wave IM (TWIM), differential mobility analyzer (DMA), and trapped IM (TIM).[125] Several research groups have used IM-MS measurements to build large, high-quality CCS libraries[242,243,252–254] which can be used as molecular descriptors for identifying unknowns or as restraints to generate three-dimensional structure models.[255,256] While such CCS databases add significant utility to IM-MS, harmonizing

such measurements acquired using different IM methods remains a challenge.[253] For example, while DTIM utilizes a well-known linear relationship between the separation field and td to extract CCS[128], similar TWIM conversions currently require the use of a power-law function that is less robust.[257] Previous reports have detailed many challenges associated with TWIM CCS calibration, which include the optimal selection of both calibrants and experimental conditions to achieve accurate TWIM CCS values.[257]

TWIM, since its introduction in 2004[147] and commercialization in 2006[148], has been one of the most widely used IM platforms.[125] TWIM utilizes a series of DC waves in a gas-filled stacked ring ion guide to separate ions. Higher mobility ion species are overtaken by the waves less often than lower mobility species and thus travel faster, thus enabling a mobility-based separation.[151] Although TWIM-MS provides a useful platform for rapid gas-phase separations, the understanding surrounding the theoretical details of ion transport necessary to robustly convert TWIM arrival times to CCS values is currently lacking. Despite this, many studies have shown experimentally that average TWIM ion velocities correlate well with CCS values from DTIM measurements.[257] TWIM theory has developed steadily since the initial introduction of the technology, beginning with expressions describing average ion velocity and resolving power,[152] with more recent additions including advancements in computational frameworks for TWIM CCS calibration[199] and empirical predictive expressions for TWIM peak width.[245] Most recently, a physics-based model of TWIM ion transport was developed incorporating velocity relaxation and TW anharmonicity.[237]

In this report, we describe a new TWIM calibration method based in-part on recent advancements in TWIM theory,[237] that yields TWIM CCS values across a significantly broader range of analyte molecules and TW conditions with dramatically improved accuracies, precisions and robustness when compared to previous power-law calibration procedures. In addition, our new

calibration procedure obviates the need to carefully select calibrant ions and instrument conditions, supports TW amplitude as well as velocity-ramped data, and performs well on the recently described Cyclic TWIM instrument configuration. We benchmarked our new calibration workflow using a variety of molecules including small molecules, lipids, peptides, proteins, and protein complexes, measured under a wide range of instrument conditions. We conclude by evaluating the current state-of-the-art for high-precision CCS measurements, and discuss the future promise and challenges that such measurements represent.

4.2 Experimental Section

4.2.1 Materials and Reagents

Ubiquitin (U6253), cytochrome c (C2506), β-lactoglobulin (L7880), avidin (A9275), bovine serum albumin (BSA) (P7656), transthyretin (P1742), concanavalin A (ConA) (C2010), alcohol dehydrogenase (ADH) (A7011), pyruvate kinase (PVK) (P9136), and glutamate dehydrogenase (GDH) (G7882) were purchased from Sigma-Aldrich, USA. For native-like ions, protein samples were prepared in 200 mM ammonium acetate and sprayed at a final concentration of 5 μM. For denatured protein ions, samples were prepared in methanol/water/acetic acid (49%/49%/2%) solution. D, L Polyalanine (Sigma-Aldrich, P9003) was prepared at a concentration of 1mg/ml in acetonitrile/water/acetic acid (49.5%/49.5%/1%) solution. Reverse peptides SDGRG (Sigma-Aldrich, S3771) and GRGDS (Sigma-Aldrich, G4391) were prepared in water/methanol/acetic acid (49%/49%/2%) solution. Egg sphingomyelin (SM) (860061C), 18:1 (Δ9-Cis) 1, 2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) (850725C), 14:0 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) (850345C), and 16:0-18:1 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (850457C) were purchased from Avanti Polar

Lipids, USA. Lipid samples were prepared in chloroform/methanol/formic acid (90%/9.9%/0.1%) solution to a final concentration of 10 μg/ml. Glutathione (Sigma-Aldrich, G4251), l-lysine (Sigma-Aldrich, L5501), acetaminophen (Sigma-Aldrich, A7085), and caffeine (Sigma-Aldrich, C53) were provided by the Kennedy group at University of Michigan and prepared at 1 mM final concertation in water/formic acid (99%/1%) solution.

4.2.2 Instrumentation

TWIM ATDs were collected on Waters Synapt G2 IM-MS and Waters Select Series Cyclic IM-MS instruments. Instrumentation details can be found elsewhere[151,247]. Briefly, ions were generated using nano electrospray ionization (nESI), and then pulsed into the TWIM cell. For the Synapt G2, voltages throughout the instrument were optimized to transmit native protein ions without significant activation. TWIM cell was operated at a pressure of ~3.4 mbar (200 ml/min and 90 ml/min flow rates for He and N_2, respectively) and ATDs were collected at a range of wave height and wave velocity conditions. For cyclic IM-MS, data were collected using a single pass mobility separation using an IM pressure of ~ 2mbar N_2.

4.2.3 CCS Calibration and Data Processing

TWIM ATDs from Synapt G2 were extracted using TWIMExtract[206]. ATDs were first transformed using a one-dimensional Gaussian filter (using Scipy[204,205]) to convert all signals to an IM resolving power of ~10 in order to match reference data. These modified ATDs were then fit using a Gaussian function to determine their centroids. Ion arrival times were corrected for ion travel within the time of flight mass analyzer as described previously.[244] These values served as the inputs for CCS calibrations.

We performed TWIM CCS calibrations using the following functions: 1) power law, 2) power law + radial correction, 3) blend, and 4) blend + radial correction (Refer to the **Theory** section below

for detailed definitions for each function). Calibrations using functions 1 and 2 were carried out using custom python script using modules from numpy and scipy, whereas calibrations using equations 3 and 4 were carried out using IMSCal written in C. IMSCal uses a Bayesian inference method to predict CCS values with prior distributions pre-determined based on some of the data contained in this report. Reference CCS values for calibrations discussed below were taken from various published databases.[243,253] Leave-one-out cross-validation was carried out by creating calibration sets and eliminating data associated with one chemical species (including all relevant charge states) and using this modified dataset as input for CCS predictions. All calibrations were performed in an automated manner using custom python scripts. All calibrated CCS data were analyzed using custom python scripts using Numpy and Scipy libraries.[204,205] Matplotlib[207] was used to generate all the data figures shown in this work.

4.3 Theory and Modelling of Travelling Wave Ion Mobility

In an ideal experiment, the measured transit time of an ion through a TWIM device would be sufficient to determine its ion mobility and, consequently, its CCS. This is difficult in practice because the three dimensional motion of ions in typical TWIM devices is complicated, whereas practical implementations of TWIM currently utilize waves that move forward in steps rather than smoothly, and the temperature and pressure of TWIM separators are not usually known. The established method of calibration[155] of TWIM data utilizes an empirically determined power-law relationship, equivalent to the following expression:

$$\overline{v_{ion}} = \overline{v_{power}}(b, c) = bK^c \qquad 4-1$$

where $\overline{v_{ion}}$ is the measured average ion velocity in the TWIM device (corrected for any known time offsets), K is the ion mobility, and b and c are calibration parameters.

The combined theoretical and modelling approach that underpins the advances made in this report has been described in detail elsewhere.[237] Here, we give a very brief summary of the previous theory that we build upon to produce our finalized calibration functions. Ultimately, we aim to construct a calibration function which includes as much of the known physics of the device as is reasonably possible, but also allows for sufficient additional freedom to allow it to describe the motion of ions in real devices. We begin by considering an idealized one dimensional smoothly moving travelling wave within a TWIM device, having constant wave velocity and amplitude. Translating this expression into the frame of reference of the moving wave, the motion of ions in such a device can be described in dimensionless coordinates using the following differential equation of motion

$$\alpha \frac{d^2 z(\tau)}{d\tau^2} = -\frac{dz(\tau)}{d(\tau)} - 1 + \gamma sinz(\tau) \qquad 4-2$$

where $z(\tau)$ is the position of the ion as a function of the dimensionless time coordinate τ,

$$\gamma = 2\pi \frac{V_o}{v\lambda} K \qquad 4-3a$$

$$\alpha = 2\pi \frac{v}{\lambda} \frac{Km}{q} \qquad 4-3b$$

K is the ion mobility, V_0 is the on-axis wave amplitude, v is the wave velocity, λ is the wave-length, and m/q is the mass-to-charge ratio of the ion. In experiments focused on small molecules conducted under typical TWIM conditions, α is small, and the approximation $\alpha = 0$ is often appropriate. In this case, the ion velocity in the laboratory frame depends only on γ and is given by $\overline{v_{ion}} = v(1 - \omega_0)$ where $\omega_0 = \sqrt{1 - \gamma^2}$. For large molecules such as proteins, however, velocity relaxation effects corresponding to ion acceleration and deceleration cannot usually be ignored, resulting in measured td values that depend upon α and, therefore, m/q. Although eq. 4-

2 cannot be solved exactly in this case, it can be shown (extending previous results[237]) that for arbitrary γ and small α the resulting average ion velocity in the laboratory frame is, to order α^6.

This expression gives velocities matching numerical solutions of eq. 4-2 to within 0.5% for $\alpha < 0.4$. However, for large protein complexes, we often use $\alpha \sim 1$ and $\gamma \ll 1$. Therefore, we will utilize the following new perturbative result obtained up to the γ^6 order using the Lindstedt-Ponicare[258] method

$$\overline{v_{\gamma,6}} = \frac{v}{1+\alpha^2}\left[\frac{\gamma^2}{2} + \frac{\gamma^4}{4}\frac{1+10\alpha^2+15\alpha^4}{(1+\alpha^2)^2(1+4\alpha^2)} + \frac{\gamma^6}{16}\frac{1+23\alpha^2+234\alpha^4+1171\alpha^6+2291\alpha^8+1620\alpha^{10}}{(1+\alpha^2)^4(1+4\alpha^2)^2(1+9\alpha^2)}\right] \qquad 4-4$$

This expression is accurate to within 0.5% for $\gamma < 0.55$. A single expression for average ion velocity (that maintains our target accuracy for all conditions that we are likely to encounter in routine TWIM experiments) is obtained by combining these two expansions as follows

$$\overline{v_{blend}} = \frac{\gamma^n}{\gamma^n + \alpha^m}\overline{v_{\alpha,6}} + \frac{\alpha^m}{\gamma^n+\alpha^m}\overline{v_{\gamma,6}} \qquad 4-5$$

where the choice $m = 8$ and $n = 12$ maintain accuracy relative to the numerical solution.

For the reasons given above, the motion of real ions is not exactly described by eq. 4-2, and thus we have found it necessary to introduce additional freedom in eq. 4-5 to promote high-quality TWIM CCS calibration. A natural choice, given the uncertainties involved in TWIM pressures and temperatures, is simply to replace the parameters α and γ with $a\alpha$ and $g\gamma$, respectively, to give a function $\overline{v_{blend}}(a, g)$. The new calibration parameters are expected to have values close to unity, which aides numerical fitting routines.

In addition to the velocity relaxation effects described above, detailed ion optical modelling reveals that different populations of ions can take up different radial distributions in RF-confined TWIM devices. For example, singly charged peptide ions are generally less well confined than multiply charged peptide ions of the same mobility.[259] As a consequence, such ions are, on average,

closer to electrodes that produce the travelling waves and effectively experience larger E. Our simulation suggests that this effect can be accommodated by multiplying the average ion velocity by an additional correction factor to give

$$\overline{v_{blend+radial}}(a,g,d) = e^{\frac{d}{\sqrt{z}}}\overline{v_{blend}}(a,g) \qquad 4-6$$

and

$$\overline{v_{blend+power}}(b,c,d) = e^{\frac{d}{\sqrt{z}}}\overline{v_{power}}(b,c) \qquad 4-7$$

where d is an additional calibration parameter. In the remainder of the paper we investigate and compare calibrations obtained utilizing the forms $\overline{v_{power}}(b,c)$,(power law) $\overline{v_{power+radial}}(b,c,d)$, (power law + radial) $\overline{v_{blend}}(a,g)$, (blend), and $\overline{v_{blend+radial}}(a,g,d)$ (blend + radial)

4.4 Results and Discussion

To test the four TWIM CCS calibration functions above, we performed large scale calibrations using a pool of analyte ions containing native-like proteins, denatured proteins, peptides, and small molecules and metabolites (Figure 4-1A). The distribution of ions and their CCS values are shown in Figure 4-1A. Also, we collected TWIM data across a wide range of wave height (WH) and wave velocity (WV) conditions. We first examined the calibration error as the per cent root mean square error in CCS (% CCS RMSE) by including all the ions in the calibrant set. TWIM CCS calibrations utilizing the power law function produced the greatest % CCS RMSE values across most WV values (Figure 4-1B). For example, we observe a maximum error of 8.15 % at 1000 m/s and 35 V, while achieving a minimum error of 1.39 % at 400 m/s and 20 V, similar to previous reports.[260] Adding a radial correction term to the power law TWIM CCS calibration

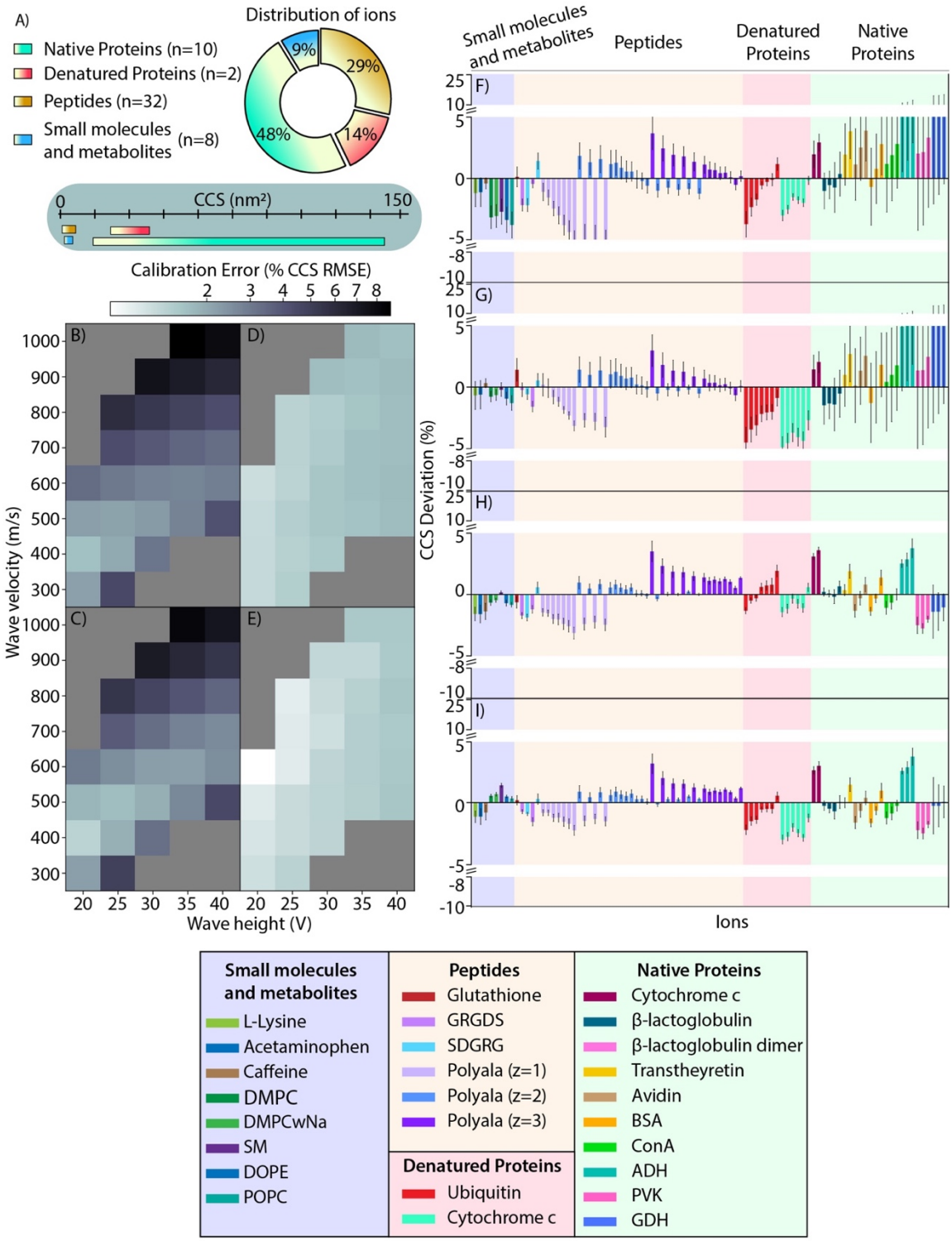


Figure 4-1. Performance of the calibration methods over a matrix of TWIM conditions for all ions. A) Overview of the included molecular classes. The head maps show the dependence of % RMSE calibration error on TWIM conditions using calibration expressions B) power law, C) power law + radial, D) blend, and E) blend + radial term. In the 'blend' calibration, the parameter 'a' was allowed to vary. Plots F-I) show the corresponding CCS deviations obtained by leave one species out cross-validation. The solid bars are averages over all TWIM conditions and the error bars show one standard deviation.

function does not significantly improve the errors encountered, as evidenced by the similar trend

in % CCS RMSE observed in Figure 4-1C. Next, we evaluated the CCS calibration errors associated with using the blend function (Figure 4-1D), where we achieve significantly reduced % CCS RMSE values when compared with power law functions. We do not detect any WV dependency in calibration accuracy using this function, which yields an average % CCS RMSE value of 1.60 ± 0.16 %, significantly lower than that achieved for power law functions. The blend + radial function performs similarly to blend function, but exhibits an improved % CCS RMSE value of 1.50 ± 0.16 % (Figure 4-1D). Taken together, our data supports significantly improved CCS precisions and accuracies can be extracted from blend-based functions when compared with those based on the power law function.

We continued to evaluate and compare blend and power-law calibration functions by performing cross-validation tests of each across all TWIM conditions in an effort to assess their robustness. In our cross-validation approach, we iteratively left one species out of the calibrant set and treated those ions as "unknowns". Figure 4-1F and 4-1G show the deviation in measured CCSs when compared to database values[244,253] for unknown ions using the power law and power law + radial functions, respectively. Without the radial correction term, we observe systematic negative errors for power-law calibrations associated with low molecular weight ions, and switching to large positive deviations for larger ions (Figure 4-1F). Incorporating a radial correction factor reduces the errors encountered for low molecular weight species s. For instance, all the ions measured in polyalanine series and small molecule dataset exhibit % CCS deviations of ± 4%. However, the radial term does not improve errors associated with high molecular weight species relative to those observed for the unmodified power law (Figure 4-1G). The blend and blend + radial functions both produce dramatically reduced % CCS deviations when compared to power-law functions, producing values within ± 5% (Figure 4-1H and 4-1I). In addition, no evidence of

systematic error based on ion molecular weight is observed. Both blend functions both perform similarly in a general sense, producing average % absolute CCS deviation values of 1.21 ± 0.91% and 1.27 ± 0.98%. Overall, cross-validation data suggests that blend CCS calibration functions exhibit superior robustness when compared to equivalent power-law functions.

We then examined the extent of precision, accuracy and robustness provided by TWIM CCS calibration functions for the specific molecular classes included in our overall dataset using cross-validation as above. Figure 4-2 displays the % CCS deviation values determined using all four calibration functions for each molecular class encompassed in our analysis, including: all ions, small molecules and metabolites, peptides, and native proteins. For this analysis, and the rest of the analyses in this report, we fixed the **a** parameter found in the blend and blend + radial functions so that it scaled linearly from 0.89 to 1.17 as wave velocity is increased from 300 to 1000 m/s. This scaling function was used to reduce over-fitting of TWIM data. As discussed in the context of Figure 4-1, the analysis shown in Figure 4-2 highlights the large CCS deviations produced using power law and power law + radial functions, which stands in contrast to those

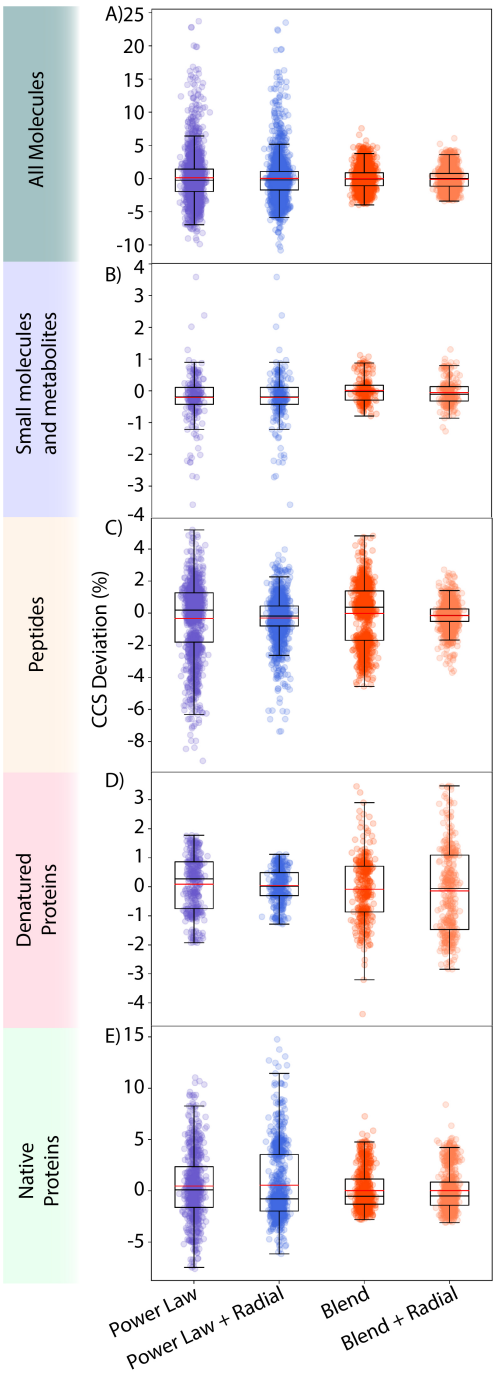


Figure 4-2. Comparison of all-molecule and molecule class-specific calibrations produced using power law and blend functions with and without the radial correction term. In both blend calibrations, the parameter 'a' was scaled linearly using a = 0.0004*WV + 0.77

observed for the blend and blend + radial calibrations performed. Specifically, the 5th and 95th percentile values of the % CCS deviation distribution for power law are -5.78 % and 8.75 %, and -5.86 % and 7.31 % when the radial term is added. In contrast, the blend function produces 5th and 95th percentile values of -2.52 % and 3.01%, respectively, with an average % CCS deviation of -0.07 ± 1.585 %. Similarly, the blend + radial function generates 5th and 95th percentile values are -2.32 % and 2.58 %, respectively, with an average % CCS deviation of -0.081 ± 1.42 %. For specific compound classes, the blend and blend + radial functions tend to predict more accurate CCS values than power law and power law + radial functions for small molecules and metabolites (Figure 4-2B), peptides (Figure 4-2C), and native proteins (Figure 4-2D). Contrastingly, the power law functions out-perform blend functions for the denatured proteins standards analyzed here, although the magnitude of the precision increase equates to an improvement of ~+/- 1% CCS RMSD. It is worth noting that our denatured protein ion set contains only two species, ubiquitin and cytochrome c. Therefore, leaving one species leaves one unknown set for analysis. In all these cases, the blend + radial function exhibits slightly better CCS precision and accuracy values than the blend function in our cross-validation analyses. This is likely due to the radial correction term included in the former function that acts to correct ion transit times for lower charge state ions.

Following our cross-validation analyses using data from linear TWIM analyzers, we applied the blend + radial function to calibrate TWIM arrival times recorded on a cyclic TWIM instrument platform. Cyclic TWIM has the potential to separate using R values in excess of 300, and thus enabling the acquisition of more accurate TWIM CCS values from complex mixtures than previously possible.[247] To evaluate our next-generation calibration functions for cyclic TWIM data, we used a set of calibrant ions containing small molecules, peptides, and native proteins to

determine % CCS errors using the cross-validation approach described above. The % CCS deviations values recorded for ions using the power law + radial and blend + radial functions are shown in figure 4-3A and 4-3B, respectively. As for linear TWIM data, we observe significant m/q dependent errors for data calibrated using the power law function (Figure 4-3A). In contrast, the blend + radial function generates % CCS deviations within ± 3 % for all ions tested.

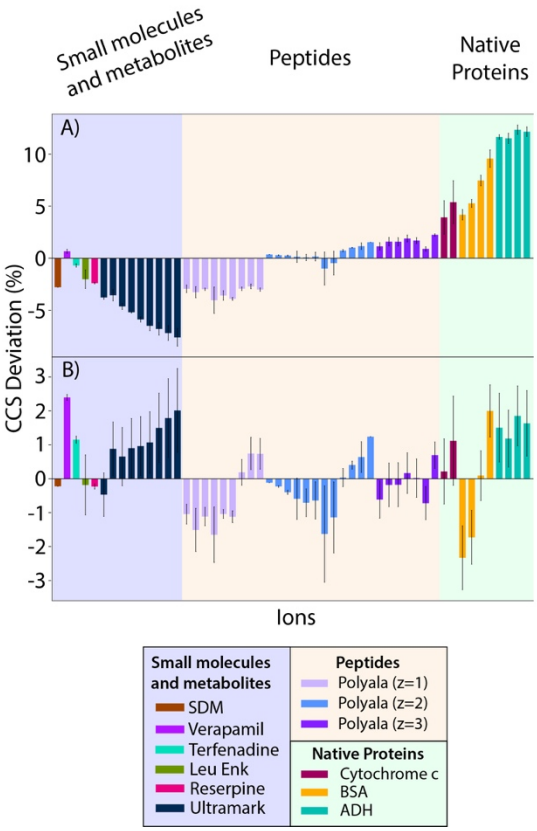


Figure 4-3. Leave one species out cross validation of all-ion calibration data obtained using cyclic TWIM instrument. A) Power law with radial term, B) blend with radial term. The parameter 'a' was fixed at a value of 1.

Typically, power law calibrations are performed using either carefully selected calibrant ions that match the molecular class of the unknowns under investigation or seek to bracket the expected K values of unknowns, in order to achieve high-accuracy TWIM CCS results. Given the data shown in Figures 4-1 to 4-3 above, we sought to investigate the ability to use a limited calibration set to perform comprehensive TWIM calibrations exhibiting high precisions and accuracies. We selected BSA (14^+, 15^+, and 16^+) and the reverse peptides SDGRG and GRGDS (1+ and 2+) to include in this limited calibration set after a survey of multiple analyte groups present within our initial calibration set used above. Figure 4-4 tracks the % CCS RMSE values determined for these limited set calibration experiments using the blend + radial function for small molecules and metabolites, peptides, denatured proteins, and native proteins across a range of WV and WH settings. The maximum % CCCS RMSE we observe in these experiments is 2.2 %, further

illustrating the robustness of the blend + radial function, which must be extrapolated to both lesser and greater CCS values to achieve the results shown in Figure 4-4.

Figure 4-4E displays the % CCS deviation values determined for all ions studied here, integrated across all WH and WV conditions probed. In this case, the blend + radial function yields an average % CCS deviation for all ions of 0.54 ± 1.37 %. Furthermore, we analyzed the % CCS deviation as a function of molecular class as shown in Figures 4-4F and 4-4G. In addition to combining all WH and WV conditions, we also examined the WH and WV condition that produced the lowest % CCS RMSE to estimate the maximum accuracy and precision of blend + radial CCS calibration results. Figure 4-4F shows the box plot displaying % CCS deviation data overlaid with individual data points recorded for condition shown. Figure 4-4G provides a similar analysis in absolute % CCS deviation space. From these data, we obtain absolute

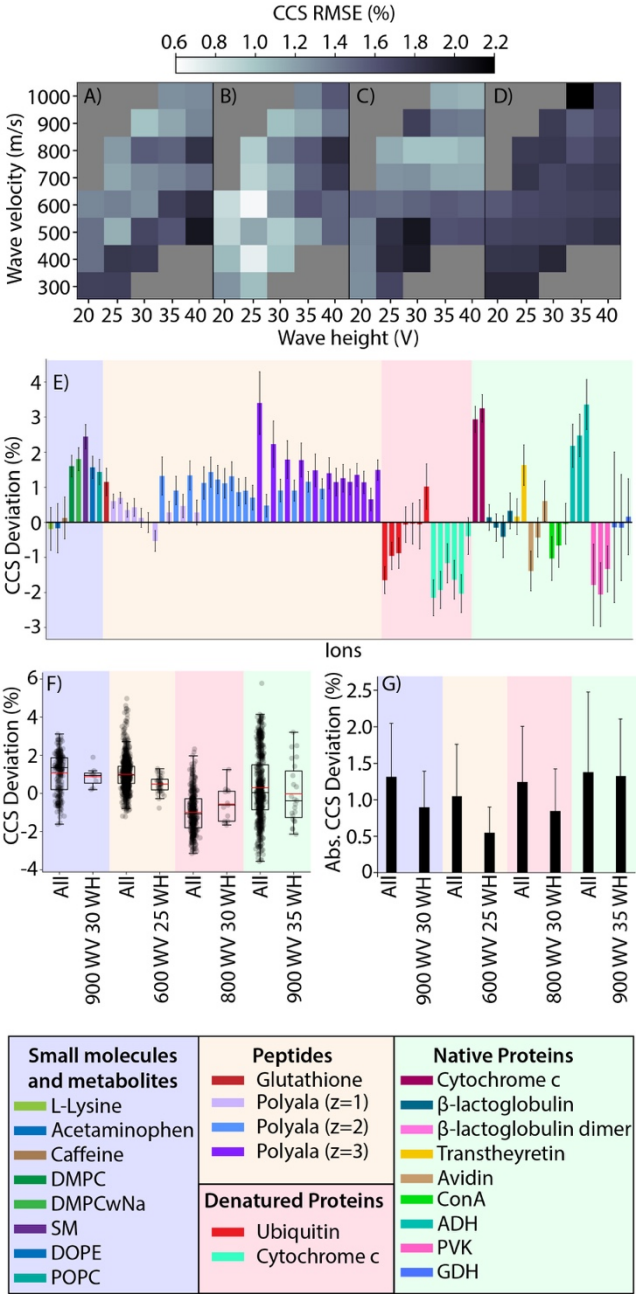


Figure 4-4. Blend with radial term calibration created using the minimal calibration set consisting of 14+ - 16+ charge states of BSA and 1+ and 2+ charge states of reverse peptides (SDGRG and GRGDS). The heat maps show % CCS RMSE for each molecular class: A) small molecules and metabolites, B) peptides, C) denature proteins, and D) native proteins. E) shows CCS deviations for individual ions. F) and G) contrast results obtained for all conditions with those obtained using optimal conditions chosen for each class. The parameter 'a' was scaled as described previously.

CCS deviation values of 0.89 ± 0.51 %, 0.54 ± 0.35 %, 0.85 ± 0.58 %, and 1.32 ± 0.79 % for small molecules and metabolites, peptides, denature proteins, and native proteins ions respectively. These results represent the lowest CCS error values reported for TWIM measurements to date. The CCS error values reported for small molecules and metabolites and peptide ions are comparable with recently-reported drift tube IM values.[261] For native protein ions, low CCS errors will clearly aid in generating accurate three-dimensional models, where such data are routinely used as restraints for MD simulations.[145,193]

4.5 Conclusions

In this study, we introduce a series of new calibration functions for TWIM CCS calibration. We devised these functions using TWIM theory to properly capture ion trajectories during TWIM separation. Also, a radial term was introduced that corrected for low charged ions that experience higher fields in TWIM. Thorough validation and testing of these new functions against a large range of analyte ions comprising of small molecules and metabolites, peptides, denatured proteins and protein complexes, across a wide range of WH and WV conditions, provides a comprehensive picture of current TWIM CCS determination capabilities. Our results demonstrate that the blend and blend + radial functions outperform power-law calibrations in almost all cases. Using cross-validation, we determined that the blend and blend + radial functions produce smaller % CCS deviations for all ions than CCS calibrations carried out using power law functions. The blend + radial function produced results of superior accuracy, precision and robustness in our analysis. Cyclic IM data was also successfully calibrated using blend + radial function and yielded low error CCS values. Finally, we tested the robustness of the blend + radial function by using a limited

calibration mixture containing BSA and reverse peptides producing CCS values of exceptional precision and accuracy.

Future developments in TWIM analyzer design will unquestionably push the resolving power and information content of IM measurements to higher levels. In that vein, TWIM theory must keep pace with advancements in TWIM instrumentation, in order to enable the extraction of valuable information from this forthcoming high-performance data. The procedures and data discussed here provide a concrete step forward in TWIM data analysis, enabling the extraction of CCS data with greater precision and accuracy than was possible previously. In addition, the data shown in Figure 4-4, where a limited set of calibrants is used to produce high quality CCS data, independent of ion class or K, will act to dramatically increase the throughput of TWIM CCS measurements. Finally, it is worth noting that the CCS values determined here are nearing, and in some cases exceeding, the precision of the standards used for calibration. As the IM-MS community continues to push toward results of greater accuracy, precision and resolving power, it is clear that we must re-engage in efforts to provide high-quality data for calibration standards for this rapidly evolving area.

Chapter 5 CIUSuite2 Classification: A method for high throughput classification of collision induced unfolding fingerprints

(Adapted with permission from: **1)** Polasky, D. A., Dixit, S. M., Fantin, S. M., and Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019**, *91* (4), 3147-3155 and **2)** Polasky, D. A., Dixit, S. M., Vallejo, D. D., Kulju, K. D., and Ruotolo, B. T. An Algorithm for Building Multi-State Classifiers Based on Collision-Induced Unfolding Data. *Anal. Chem.* **2019**, *91* (6), 10407-10412.)

5.1 Introduction

Native mass spectrometry (MS) has become a widespread technique in structural biology due to its ability to preserve noncovalently associated protein-protein and protein-ligand contacts and determine the stoichiometry and connectivity of these interactions.[262–264] The coupling of ion mobility to mass spectrometry (IM-MS) provides molecular shape information in addition to ion mass and charge, which has proven invaluable in interrogating complex biomolecular structures.[265,266] Native IM-MS has seen dramatic growth in recent years, with applications in biotherapeutic characterization[267] and drug discovery,[268] joining more traditional analyses of protein complex structure and stoichiometry. A significant challenge in IM-MS is the separation of closely related protein conformations, as biologically relevant conformational variations often occur beyond the resolution limits of modern IM instrumentation. However, gas-phase activation provides a powerful approach to probe these subtle structural differences by assessing the resulting pattern of intermediate structural families produced from collisional heating of gas-phase protein ions. This approach, termed collision-induced unfolding (CIU) when used with collisional

activation, has a rich history in protein structure evaluation[157,269,270] and has seen rapid growth for drug discovery[160,169,271] and biotherapeutic characterization[164,168,272–275] in recent years. The relative speed of CIU combined with detailed comparative structure information make it a promising technique for the development of structure-sensitive screening methods at medium to high throughput.

The development of screening methods using CIU has seen growing interest in recent years. A number of reports have demonstrated proof-of-principle approaches using CIU to distinguish ligand binding sites for drug discovery in kinases,[160,167,169] cooperative binding of ligands to a protein complex,[159] and contributions of allosteric effects of ligand binding on protein structure.[276] Screening approaches sensitive to these structural parameters are in great demand in drug discovery and the growing biotherapeutic development fields. The relative comparison of CIU fingerprints under different conditions, for example following ligand binding to a target protein or after applying heat stress to a biotherapeutic, enables the determination of useful information about the structure of a protein and its response to perturbations. However, many of the CIU studies discussed above utilized manual analysis for all or part of the CIU data processing, which would substantially limit the high throughput of such experiments.

To address these challenges, we have developed a classification algorithm in CIUSuite 2 utilizing robust statistical methods to enable robust classification of CIU data across a broad range of applications. CIUSuite 2 is a software package that utilizes established fitting and statistical methods to enable the robust quantitation of CIU data across a broad range of CIU applications and analysis types, especially for enabling the analysis of low intensity CIU datasets. CIU fingerprint classification is implemented using linear discriminant analysis and support vector machines to enable next-generation high throughput screening experiments. Most of the CIU

comparisons are done using a single charge state analysis. However, recent work has demonstrated the benefits of including CIU information from multiple charge states in distinguishing the structures of monoclonal antibodies.[277] The incorporation of all information available from multiple charge states provides, in principle, great potential for improving CIU classification and screening methods without increasing data acquisition time. So, in this report, we devise a classification workflow that can include data from multiple charge states, improve processing speed to enable processing of large datasets, and expand the scope to include "subclass" analyses beyond charge states. We demonstrate the utility of the classification method in CIUSuite 2 to characterize biotherapeutics and ligand binding modes in a drug discovery context.

5.2 Experimental Section

5.2.1 Sample Preparation

IgG1κ, IgG2κ, IgG3κ, and IgG4κ from human myeloma were purchased from Sigma-Aldrich (St. Louis, MO). These antibody samples with a concentration of 1 mg/ml were buffer exchanged into 100 mM ammonium acetate buffer using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA) without further purification. SiLuLite SigmaMab Universal antibody standards, IgG1λ, and IgG4λ from human myeloma were purchased from Sigma-Aldrich and supplied as lyophilized powder (St. Louis, MO). Samples were reconstituted using Milli-Q water (Millipore) to a concentration of 2 mg/mL unless specified otherwise. Avastin® (Genentech, 25 mg/mL) and Avegra® (Biocad, 25 mg/mL) were purchased and supplied in solution formulation (158.6 mM Trehalose dehydrate, 40.9mM Sodium Phosphate, 0.16% Polysorbate 80, pH 6.2). Biotherapeutic samples were diluted to 1mg/mL using 0.9% bacteriostatic sodium chloride injection, USP. (Pfizer Inc. New York City, NY). Stressed samples were incubated at 40 ºC with 250 RPM orbital shaking

for 4 weeks. All antibody samples were buffer exchanged into 200 mM ammonium acetate buffer using Micro Bio-spin 30 columns (Bio-Rad, Hercules, CA). Buffer exchanged samples were then diluted to a working concentration of 1 mg/mL (~6.7 μM).

Translocator Protein (TSPO) was purified and expressed using established protocols.[278] All lipids purchased from Avanti Polar Lipids (Alabaster, AL). Ammonium acetate and Octyl β-D-glucopyranoside (OG) were purchased from Sigma Aldrich (St. Louis, MO). TSPO was buffer and detergent exchanged simultaneously from 5 mM Tris, 150 mM NaCl , 0.20% DM, pH 8.0 to 40 mM OG , 200 mM ammonium acetate, pH 8.0 using 100kDa Amicon Ultra-0.5 Centrifugal Filter Units (MilliporeSigma, Burlington, MA). Lipid binding studies were performed following established protocols.[279]

Src kinase domain DNA was synthesized by GeneArt (Life Technologies, Grand Island, NY) using E. coli modified codons and subcloned into pET28a with a modified TEV-protease cleavable N-terminal 6x-His tag. The plasmid was transformed by electroporation into BL21 DE3 electrochemically competent cells with a YopH in pCDFDuet-1. Cell growth, protein expression, and purification were adapted from protocols previously developed for the c-Src kinase domain[280] without cleavage of the His-tag. Dasatinib, staurosporine, foretinib, and ponatinib were purchased from LC Laboratories (Woburn, MA). Protein was reconstituted and buffer exchanged into 200 mM ammonium acetate (Sigma-Aldrich, St. Louis, MO) at pH 7.0 using Micro Bio-Spin 6 columns (BioRad, Hercules, CA) to a final concentration of 10 μM. Samples were incubated at a ratio of 3:1 inhibitor:protein, on ice for 15 minutes prior to analysis by IM-MS.

5.2.2 CIU Acquisition

All CIU data were acquired using a Synapt G2 quadrupole-ion mobility-time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Waters, Milford, MA). Sample was transferred to a

gold-coated borosilicate capillary needle (prepared in-house), and ions were generated by direct infusion using a nano-electrospray ionization (nESI) in positive mode. The electrospray capillary was operated at voltages of 1.5-1.7 kV with the sampling cone at 40 V. The backing pressure was set to 7.9-8.1 mbar for antibody samples or 5.0 mbar for kinase samples. Instrument settings were tuned to completely remove the micelle prior to IM separator, including source temperature (40° C), helium cell gas flow (100 mL/min), and sampling cone (120 V). The trap collision cell was pressurized to 4-5 × 10^{-2} mbar of argon gas, helium cell flow to 1.4x10^{3} mbar, traveling-wave IM separator to 3.4 mbar, and ToF MS to 1.5 × 10^{-6} mbar. CIU experiments were performed by ramping the collision voltage in the trap cell from 5 to 200 V (antibodies), 50 to 150 V (membrane protein), or 10 to 125 V (Src kinase) in 5 V increments with a dwell time of 6 s at each collision voltage.

5.2.3 Classification

Briefly, a classification scheme is built based on training datasets from each group. These datasets can be either the raw or the Gaussian fitted CIU data. Input training data is standardized within each subclass and collision voltage by scaling to zero mean and unit variance. For input Gaussian data, each attribute of each Gaussian peak is considered separately so that centroids are only standardized with centroids, widths with widths, and so forth. Then, our workflow implements a univariate feature selection (UFS) method based on an analysis of variance (ANOVA) F-test[281] to assess the significance of activation energies capable of differentiating CIU fingerprints (Figure II-1). We iterate over all possible combinations of a training dataset in order to obtain the mean and standard deviation of $-\log_{10}$(p value) which serves as the score for each collision voltage. For multi-state classifier, standardized and labeled training data for each subclass is assessed separately by the UFS method. Next, we employ a “leave one out” cross-validation

scheme[282] that examines the accuracy of classification, which is comprised of a linear discriminant analysis[283] (LDA) step followed by support vector machine[284] (SVM) classification of the linear discriminants, using subsets of CIU data from collision voltages in decreasing order relative to the score assigned during UFS analysis. Random sampling from the possible input permutations during cross-validation was implemented to reduce the computational time for large datasets with a user-specified number of the possible permutations. Cross-validation approach enables optimal selection of collision voltages to use for the resulting model and can be used to detect under- or over-fitting in the final model selected (Figure II-2). Finally, classification is performed on the model dataset with the optimized set of collision voltages, dividing the linear discriminant space into "decision regions" corresponding to the provided groups. The resulting classification scheme can then be used to evaluate "unknown" CIU datasets (not used in training) to predict the class and probability for each unknown. We have also included a 'manual' classification mode, where users can select any number of specific collision voltages to build a classification model. This is particularly helpful in scenarios where the accuracy observed in the cross-validation step is unacceptably low.

5.3 Results and Discussion

In CIUSuite 2, we have implemented a classification workflow that uses rigorous statistical methods to generate classifying schema from known fingerprints that allows for facile evaluation of unknown samples against these schemas for rapid sorting. Data shown in Figure 5-1 displays an example implementation of our classification workflow, utilizing CIU data for immunoglobulin G (IgG) standards acquired across IgG1, IgG2, IgG3, and IgG4 subclasses. Each of our IgG CIU datasets contained four replicates, which we subdivided evenly into both training and test data in

order to evaluate our approach. Of the forty collision voltages acquired for each CIU dataset, only a few were found to be highly differentiating between classes by UFS, with 85 V having the maximum score (Figure II-3E). This voltage is near the value required for the first CIU transition for each IgG subclass (Figures II-3A – II-3D). Cross-validation of UFS results revealed a classification accuracy 92.2% using only the CIU data isolated at 85 V, and decreases as additional CIU data is added (Figure II-3F). Thus, our algorithm selected CIU data acquired at 85 V automatically in order to build a classification scheme. Figure 5-1A shows the three-dimensional plot of linear discriminants (LDs) constructed using this data, which groups IgG CIU data into well-separated clusters. Furthermore, test data clustered correctly in all cases using this classification scheme (pentagons, Figure 5-1A). We further used CIUSuite 2 to compute the probability of each test dataset clustering into each IgG subclass (Figure 5-1B), finding that each dataset was classified correctly with probability values ranging from 0.52 – 0.73 (Table II-1). In general, our CIU classification workflow is generalizable, rapidly processing data in an automated fashion and accommodating any grouping scheme.

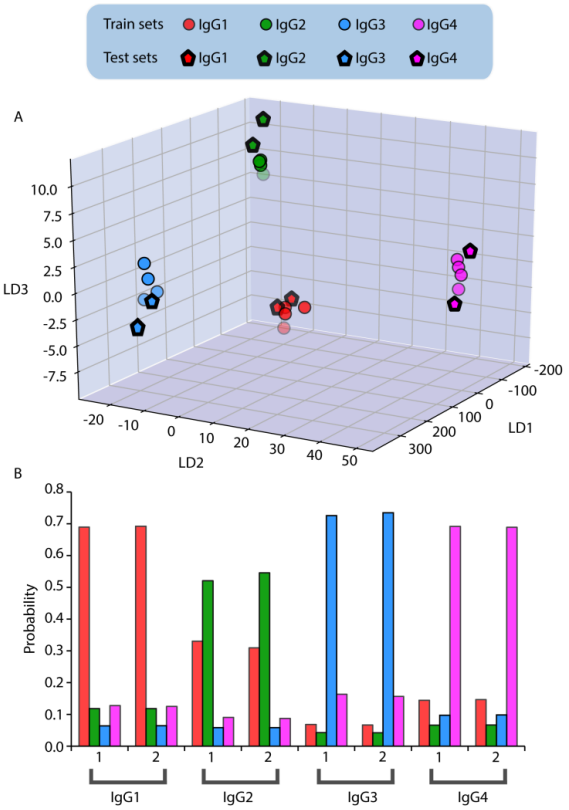


Figure 5-1. Classification of different IgG standards. (A) LDs for both training (filled circle) and test (filled pentagon) datasets corresponding to IgG1 (red), IgG2 (green), IgG3 (blue), and IgG4 (magenta) subclasses are well separated into clusters in three-dimensional space, defined by LD1, LD2, and LD3 axes. (B) Probabilities associated with each replicate (labeled as 1 and 2) in terms of categorizing the CIU data into different IgG groups. Each dataset is correctly assigned to its respective IgG subclass.

CIU has been used to characterize lipid binding to membrane proteins in order to assess stability shifts in the resulting complexes.[181,285,286] Such data have been further used to distinguish between biologically relevant and nonspecifically associated lipids in membrane proteins.[163] Counterintuitively, such assessments are often more straightforward to perform for larger proteins and complexes, as they appear at m/z and IM drift times that are frequently less contaminated by chemical noise. Since TSPO is a relatively small membrane protein complex, it is an exceptionally difficult target for CIU analysis. Preliminary screening of TSPO-lipid complexes revealed certain lipids, such as phosphatidylglycerol (PG), that significantly stabilize the protein so that CIU transitions appear distorted relative to apo protein data, making the extraction of CIU50 values even more challenging (Figure 5-2A).[287]

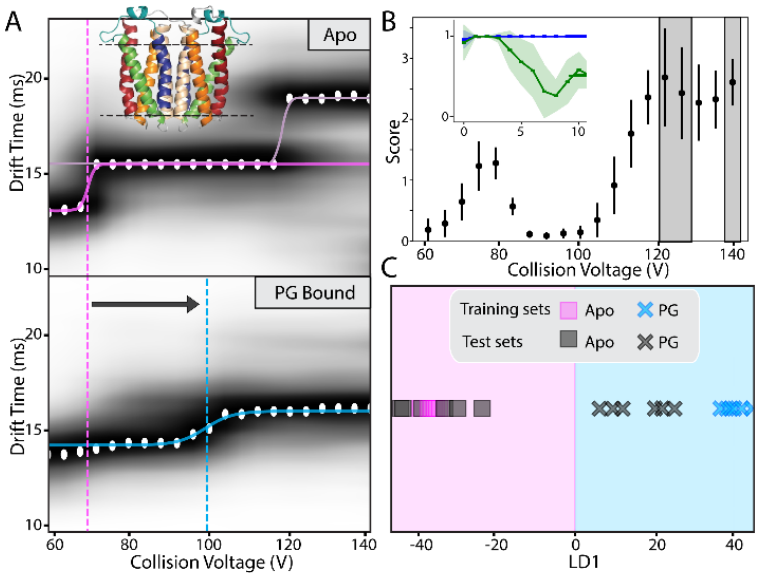


Figure 5-2. CIU50 analysis and classification TSPO-lipid complexes. (A) Feature fitting ignores low abundance chemical noise and CIU50 analysis reveals a stability shift associated with PG-bound TSPO. (B) Three voltages (120 V, 125 V, and 140 V) are used to construct a classification scheme from apo and PG-bound TSPO training sets, the inset shows a cross-validation plot indicating a high accuracy classification. (C) Additional test data sets are correctly assigned to apo (pink) or PG-bound (blue) TSPO.

To that end, we classified PG-bound and apo TSPO CIU signatures using CIUSuite 2. By using three replicates each of apo and lipid bound data to build the classification scheme, we identified 120 V, 125 V, and 140 V as the most differentiating collision voltage values in our dataset (Figure 5-2B). For validation of our classification scheme, four data sets that were not part of the training dataset were input as unknowns, and all

were correctly classified (Figure 5-2C). While it is clear that mass analysis alone could be used to identify PG bound and apo TSPO, these results illustrate a classification outcome that is exceptionally challenging to achieve using current CIU analysis tools.

Each charge state observed in a native IM-MS experiment undergoes a unique unfolding trajectory during CIU, providing potentially complementary information for a CIU-based classifier. To evaluate the utility of combining data from multiple charge states for CIU classification, we compared monoclonal antibodies IgG1 and IgG4, which differ only slightly in disulfide bonding pattern (Figure 5-3A). The native mass spectrum of IgG1 shows charge states from 22-26^{+}, with 24^{+} being the most abundant (Figure 5-3B). The CIU fingerprints for 24^{+} charge states of IgG1 and IgG4 are quite similar, aside from minor differences in the second CIU feature in the 60-80 V activation range (Figure 5-3A, bottom). Performing a single charge state comparison using the 24^{+}

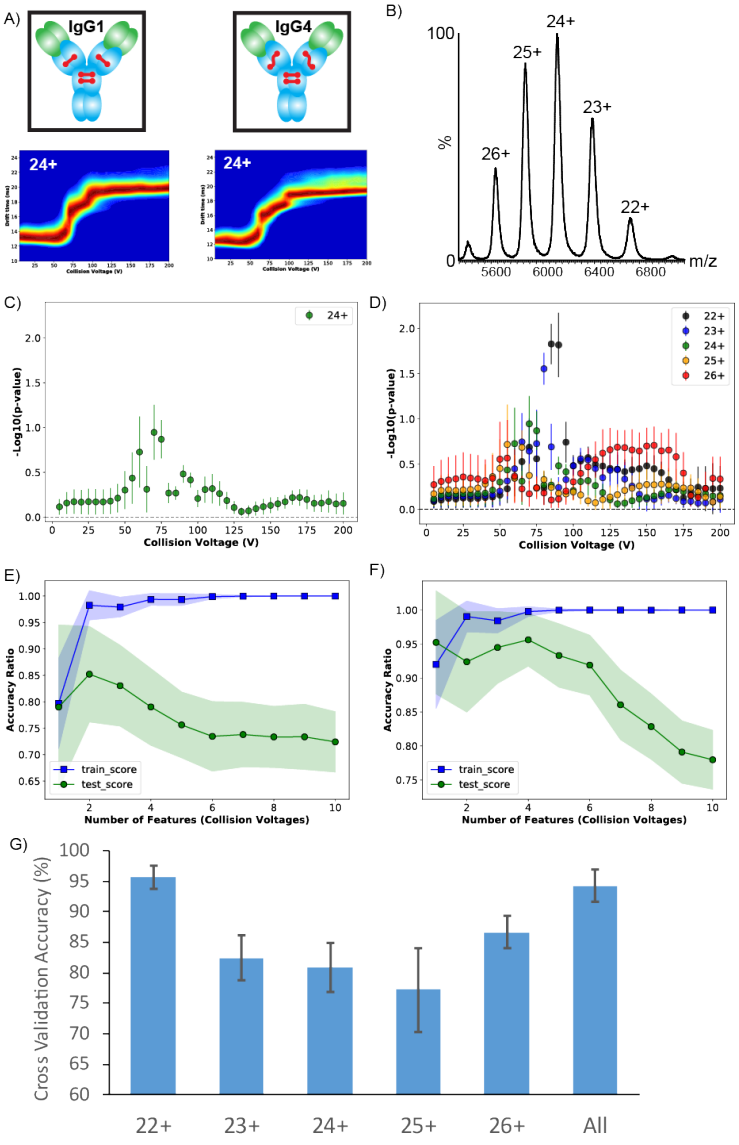


Figure 5-3. Multiple charge state classification of IgGs. A) IgG1 and IgG4 subtypes differ primarily in disulfide bond linkage, resulting in slightly different CIU fingerprints. B) Native mass spectrum of IgG1 with 22-26^{+} charge states. C) UFS score plot distinguishing IgG1 and IgG4 at the 24^{+} charge state only. D) UFS plot for all charge states of IgG1 and IgG4. E) Cross validation accuracies for 1-10 features from the 24^{+} charge state alone and F) for all charge states incorporated into one classifier. G) Optimal cross validation accuracy from each individual charge state and all charge states combined.

charge state only, as would be done in the original CIUSuite 2 workflow, results in a feature selection plot showing minor differences in the 60-80 V region as expected, with minimal difference outside that region (Figure 5-3C). The classifier that can be trained from this data is of low quality, achieving a maximum cross validation accuracy of 85% when using two features (70 and 75 V) (Figure 5-3E). Assessing all charge states with the subclass classification workflow presented here, however, reveals that the 24^+ charge state, despite being the highest signal in the mass spectrum, is not the optimal CIU trajectory to differentiate these two antibodies. To examine all charge states, we perform feature selection sequentially for each, meaning that the 22^+ charge state of IgG1 is compared to the 22^+ of IgG4, the 23^+ to 23^+, and so on. This results in five feature selection plots, which can be overlaid to evaluate each charge state (Figure 5-3D). The 22^+ charge state has the two highest scoring individual voltages (black, 85 and 90 V), followed by 80 V in the 23^+ charge state (blue), then 75 and 70 V in the 24^+ charge state (green) (Figure 5-3F). As in the single charge state classification, cross validation is performed by incorporating the data into classifiers in decreasing order of feature selection score; in this case, the input data can originate in any of the charge states rather than just a single one. The cross validation indicates that the optimal classifier in this case uses four collision voltages, two from the 22^+ charge state and one each from the 23^+ and 24^+ charge states, to achieve an accuracy of 95%, significantly improved over the 85% accuracy achieved by the classifier using just the 24^+ charge state.

To complete the comparison, we generated single charge state classifiers for all five charge states and compared the cross validation accuracy at the optimal number of collision voltages for each classifier (Figure 5-3G). Given the pair of very high scores from the 22^+ charge state in the feature selection, it is not surprising that it results in the best single charge state classifier, and indeed achieves nearly identical accuracy to the combined classifier that considered all charge

states (95%). The 23-26^{+} charge states each individually achieve accuracies in the 85-90% range, lower than the 22^{+} or combined classifiers. In this case, because one charge state is substantially better at differentiating the classes than the other charge states, its data drives the performance of the combined classifier, resulting in very similar output accuracies. Performing the classification with all charge states is, in this case, primarily a means to rapidly identify the optimal charge state and ensure it is incorporated into the final classifier. Indeed, 22^{+} is the lowest intensity charge state included in the analysis, and would not be an obvious choice from MS data alone. In cases where several charge states achieve similar feature selection scores, however, combining data from multiple charge states can generate a superior classifier to any individual charge state.

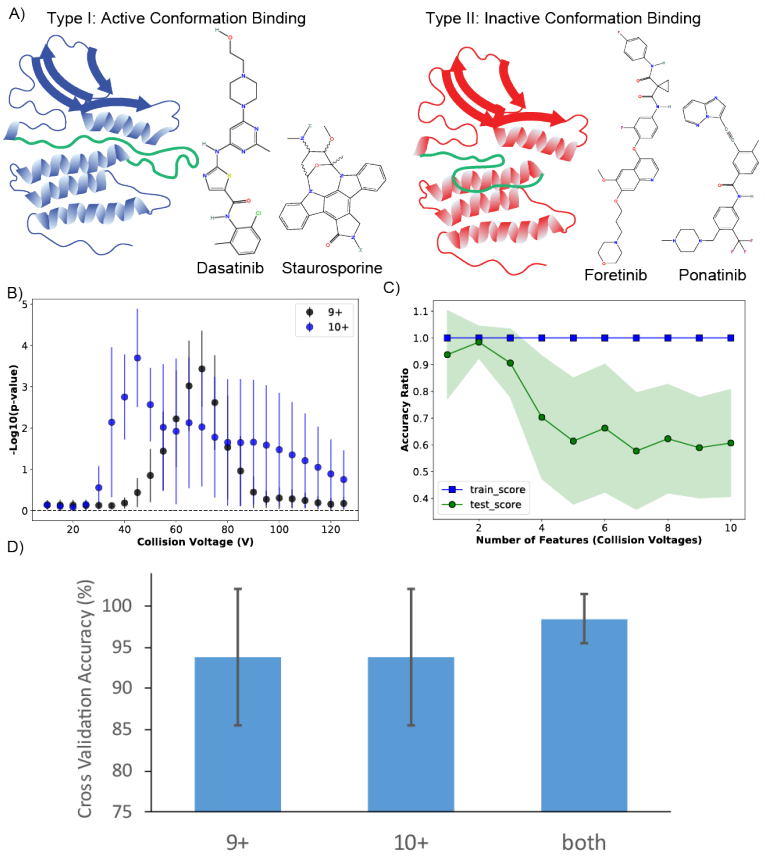


Figure 5-4. Multiple charge state classification of Src kinase. A) Type I and II kinase inhibitors target the active (left) or inactive (right) conformations of the kinase. B) UFS plot comparing Src CIU fingerprints with bound Type I (Dasatinib and Staurosporine) against Type II (Foretinib and Ponatinib) inhibitors at all charge states. C) Cross validation accuracy as a function of number of features included in the classifier for the combined classifier. D) Optimal cross validation accuracy for individual charge state and combined classifiers.

We applied this subclass classification workflow to a number of challenging applications that had previously confounded CIU classification efforts using a single charge state. Src, a non-receptor protein tyrosine kinase, plays a key role in several cell signaling processes[288,289] and has been observed to be overexpressed in

certain carcinomas and glioblastomas.[290] Several classes of inhibitors to kinases like Src are known to target different conformations of the kinase. Type I inhibitors like Dasatinib and Staurosporine bind to the active state, in which the DFG loop is in the "in" conformation, wrapping around the helices (green loop, Figure 5-4A, left). Type II inhibitors like Foretinib and Ponatinib bind the protein in the inactive conformation, in which the DFG loop is in the "out" conformation (green loop, Figure 5-4A, right). Differentiating these binding modes with the previous CIUSuite 2 classification method at a single charge state had proven challenging. Using the all charge state workflow developed here, we observe similar feature scores to distinguish the Type I and Type II inhibitors for both the 9^+ and 10^+ charge states (Figure 5-4B). As a result, the optimal classifier uses a single collision voltage each from 9^+ and 10^+, resulting in a cross validation accuracy of 98% (Figure 5-4C). The optimal classifiers for the 9^+ and 10^+ charge states individually utilized only the highest scoring single voltage in case, but achieved accuracies of only 92-93% (Figure 5-4D). The large error bars for the individual charge state classifier accuracies also indicate substantial uncertainty in their performance, with lower accuracy possible for external validation. Thus, the combined classifier using multiple charge states is superior in this case to any of the individual charge state classifiers, and enabled robust classification of ligand binding modes in a system that had proven challenging to classify with a single charge state alone.

Finally, we examined a biotherapeutic innovator/biosimilar pair, Avastin and Avegra, incorporating both multiple charge states and stress conditions into a multi-subclass classifier. Assessing the similarity of a biosimilar or generic form of an innovator biologic presents significant analytical challenges due to the size and complexity of, for example, a monoclonal antibody. Comparing higher order structure in particular is challenging using existing methods without resorting to low-throughput, atomic resolution structural biology techniques. As

biosimilars, Avastin and Avegra are highly similar and classification using all charge states of the antibodies resulted in a low degree of differentiation. The feature selection scores were low (Figure 5-5B), and despite some minor regions of difference, the optimal classifier achieved accuracy of only 87% (Figure 5-5C). Charge states are not the only subclass that can be examined with the workflow presented here, however. A subclass can be any perturbation that results in a different CIU trajectory, so long as it can be applied equally across the classes being compared. A key attribute monitored in biotherapeutics is the propensity to aggregate during transport and storage, which can be challenging to assess in the laboratory. Early warning methods for aggregation that detect structural changes following various types of stress (for example, heat or oxidation) are thus highly useful (Figure 5-5A). Avastin and Avegra were stressed by heating to 40 ºC and applying orbital shaking at 250 RPM for 4 weeks. CIU data from the stressed samples was incorporated into a classifier, along with all the observed charge

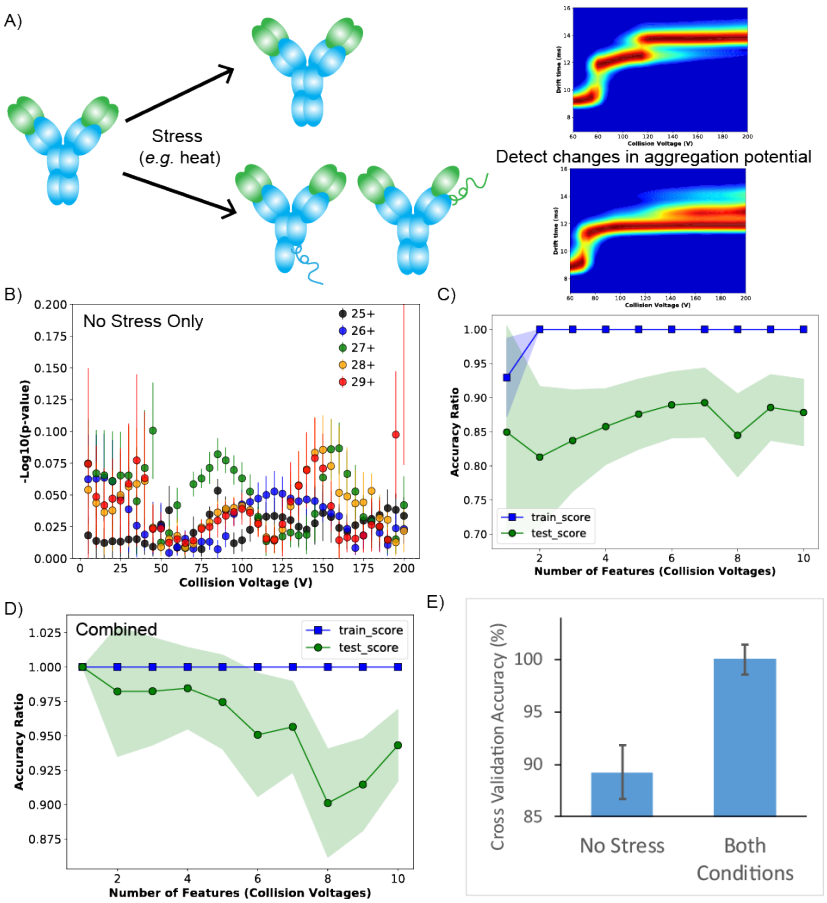


Figure 5-5. Stress subclasses distinguish Avastin and Avegra. A) Stressed antibodies can undergo structural transitions that increase their propensity to aggregate, potentially resulting in differences that can be identified using CIU. B) UFS plot for all charge states of unstressed Avastin and Avegra showing minimal differences. C) Cross validation of unstressed Avastin and Avegra. D) Cross validation from a combined classifier using stressed and unstressed Avastin and Avegra. E) Best cross validation accuracy for unstressed alone and both conditions (stressed and unstressed) Avastin and Avegra.

states, for a total of 10 subclasses (5 charge states each from stressed and unstressed conditions). The combined classifier achieved cross validation accuracy above 99% (Figure 5-5D), indicating very robust differentiation between Avastin and Avegra, significantly outperforming the classifier the used all charge states but only compared the unstressed antibodies (Figure 5-5E). Our analysis indicates that Avastin and Avegra have different structural responses to the stress employed in this study, which can be utilized to develop a classifier capable of robustly distinguishing between them using our subclass classification method.

5.4 Conclusion

CIU experiments generate rich datasets that have proven capable of distinguishing subtle differences in protein structures. This report demonstrates the utility of classification method in CIUSuite 2 that enables high-throughput screening for applications ranging from biotherapeutics characterization, lipid binding in membrane proteins, and protein-ligand interactions. Applying the subclass classification workflow presented here to analyze all charge states observed in a CIU experiment maximizes the detection of these subtle differences by incorporating more of the experimental data into the statistical framework for classification. This work indicates the potential of classification workflow to be used with a wide range of conditions or perturbations, as any change that causes differences in CIU trajectory for an analyte of interest can be incorporated into a CIU-based classifier. Incorporating differential responses to stimulus into CIU classification has the potential to make CIU sensitive to even more subtle structural differences and provide a rapid and informative workflow for evaluating protein structures.

Chapter 6 Conclusions and Future Directions

6.1 Conclusion

Studying the structures of proteins and protein complexes are tremendously important in understanding disease mechanisms and developing therapeutics. Because high resolution structural determination techniques may not always be feasible due to the complexity involved in preparing samples, and cost and time required, low resolution experiments are employed that can provide useful insights in structure and help generate accurate three-dimensional models. To this light, this thesis demonstrates the advancements made in the gas phase native MS method for studying structures of proteins and protein complexes. In particular, the development of theoretical frameworks in TWIM alongside the versatile classification workflow for CIU presented in this thesis will facilitate rapid, robust, and accurate measurements promising the utility of IM-MS technique in high throughput biophysical characterization of protein structures.

Although there are plenty of reports that use CCS information from IM-MS measurement to study the structures of proteins, there is a relatively limited number of studies that utilize the IM arrival time distributions (ATDs) to define the size the of the structural ensemble of proteins. This is due to the challenge in correlating ATDs and protein dynamics in TWIM analyzers. In chapter 2, we described a semi-empirical relationship that can provide a conformational broadening parameter reporting on the size of the structural ensemble sampled in our experiment. Additionally, our theoretical construct was able to faithfully model the ATDs across a range of TWIM conditions. We tested the semi-empirical formalism by examining the dynamics of a set of model

peptides and comparing the results with the models generated from MD simulation. Furthermore, we validated our workflow by introducing cross-links in a protein complex to reduce the structural flexibility, which was reflective on the conformational broadening parameter. The ease of the implementation of this workflow to obtain the size of structural ensemble of proteins will make TWIM MS measurements highly valuable for studying dynamic proteins, effect of ligand binding on protein structural ensemble, and intrinsically disordered proteins.

Despite the prevalence of TWIM platform in areas spanning omics to structural biology, there was a gap in the fundamental understanding of the technology. To fill that gap, in chapters 3 and 4, we have put effort to describe a novel pseudo ion trapping phenomenon and generate a new CCS calibration method in TWIM. We discovered pseudo ion trapping in TWIM by noticing aberrant resolving power values at certain TWIM conditions. This was described using a theoretical model and ion trajectory simulations illustrating that imperfect traveling waves could cause a repetitive ion motion pattern leading to ions with even small mobility difference to travel with the same mean velocity in the device. This, effectively, creates a more concentrated and compact packet of ions resulting in artificial increase in the TWIM resolving power. However, the transit times of the ions are altered when they undergo pseudo trapping in TWIM severely affecting the calibrated CCS values. Thus, our study proposed TWIM conditions to avoid pseudo trapping phenomenon for generating reliable TWIM measurements. Furthermore, we have generated new calibration functions for TWIM that combined the fundamental understanding of ion motion in TWIM comprising of ideas such as velocity relaxation and traveling wave edge effects. The new calibration functions, termed blend and blend + radial, outperform the current calibration function and yield very accurate and precise CCS values. We benchmarked the accuracy and precision of these new functions using a large scale of analyte ions comprised of small molecules and

metabolites, peptides, denature proteins, and native proteins, and across a range of traveling wave conditions. With the robustness and extrapolation capability provided by blend + radial, we proposed a reduced calibration set that can be used to obtain CCS values from TWIM measurements for any type of molecules. Therefore, our effort in advancing TWIM theory has benefited the IM-MS community to utilize the platform providing reliable and accurate measurements.

Finally, as CIU has frequently been used as a gas phase calorimetry assay, we developed a classification workflow for CIU datasets enabling high throughput characterization of protein structure and screening of protein-ligand interactions. In chapter 4, we discuss our algorithms for a single state and multi state classification of CIU fingerprints, where states can be defined as charge states of the ions, protein incubation properties, etc. We demonstrate that our method can properly classify different states usually creating a powerful model with high accuracy. Using this method, we were able to identify the class of an unknown endogenous lipid in TSPO dimer. Similarly, we were able to produce better predictive model using multi-state rather than single state classification, and this was demonstrated using Src-kinase ligand binding experiments and biotherapeutic innovator and biosimilar comparisons.

Overall, this thesis highlights the powerful IM-MS technique and its use for biophysical and structural studies of proteins and protein complexes. Additionally, valuable fundamental insights into TWIM described in this thesis will advance the field of IM-MS even more.

6.2 Future Directions

6.2.1 Direct CCS measurement from TWIM

As described in this thesis, TWIM CCS values are obtained using calibration methods. Although the calibration workflow presented in this thesis produces accurate and precise CCS values, there is still a question regarding what is the average mobility of the ion in TWIM platform? Pursuing such question will lead to even better understanding of ion trajectory in TWIM in conjunction with providing a robust theoretical framework that relates the ion's transit time directly to its mobility similar to drift tube IM analyzers. The study by Keith et al[237] has provided a differential equation describing the motion of the ion in TWIM with velocity relaxation incorporated. However, determination of the ion mobility has not been successful yet. With this thesis, we discovered the traveling wave edge effect for which we introduced a radial correction term in our calibration functions. Implementing such effect in the theory will help in constructing new models that is able to determine the average ion mobility from TWIM ATDs. Currently, the reference CCS values used for TWIM calibration are taken from a modified TWIM-MS instrument by replacing the TWIM cell with RF-confined drift tube cell. So, how does an ion's mobility from TWIM compares to the one determined from drift tube ion mobility using Mason Schamp equation? With the aim of producing large harmonious CCS database that can be used the community, answering the questions described above along with utilizing TWIM for creation of such databases will facilitate the process.

6.2.2 Next-gen theoretical CCS prediction

CCS has been used as a restraint for generating three dimensional models of proteins. However, CCS calculations can prove challenging at times due to either less accuracy or high computation time. Projection approximation (PA) is usually the fastest method but always

underestimates the CCS. Trajectory method (TM) calculations use physics-based model to determine accurate CCS values but are quite slow. To address this, IMPACT, which has a built-in model based on the correlation between PA and TM CCS values for conversion, was created as a fast CCS calculator capable of predicting TM CCS values. In chapter1 and Appendices III and IV, IMPACT was used for all theoretical CCS calculations. However, we used TM calculations on a small subset of structures leading to introducing a correction factor to convert IMPACT TM to TM calculated CCS values. TM would prove impractical to calculate theoretical CCS on tens of thousands of protein structures, which is usually the case when analyzing MD simulation output structures. Furthermore, while most IM instruments uses N_2 gas for separation and yield N_2 CCS values, IMPACT only calculates He CCS making it impractical to compare with experimental data. Due to the polarizability of N_2 molecule, TM calculation includes electrostatic interactions to model the collisions of gas molecules with the charged protein structure. This adds to the computational time as well. Therefore, there is a need for a new CCS calculation tool that is fast and produces accurate He and N_2 CCS values.

Protein's structure (Figure 6-1) encodes for its CCS. So, we can devise a tool that examines a protein's architecture and predicts the CCS values. We can employ a machine learning method trained on large dataset of protein structures of various different folds and include features such as the length of the protein sequence, three-dimensional distance map, charge on the residues, and the TM calculated CCS values in both N_2 and He. Regression models such as logistic regression and Gaussian process regression, and Bayesian models can be used for machine learning. Thorough examination of features and proper validations will be required to evaluate the

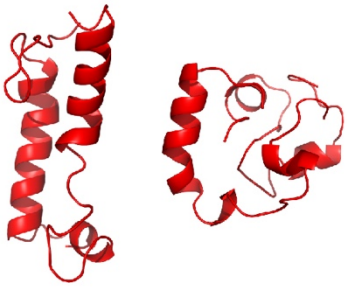
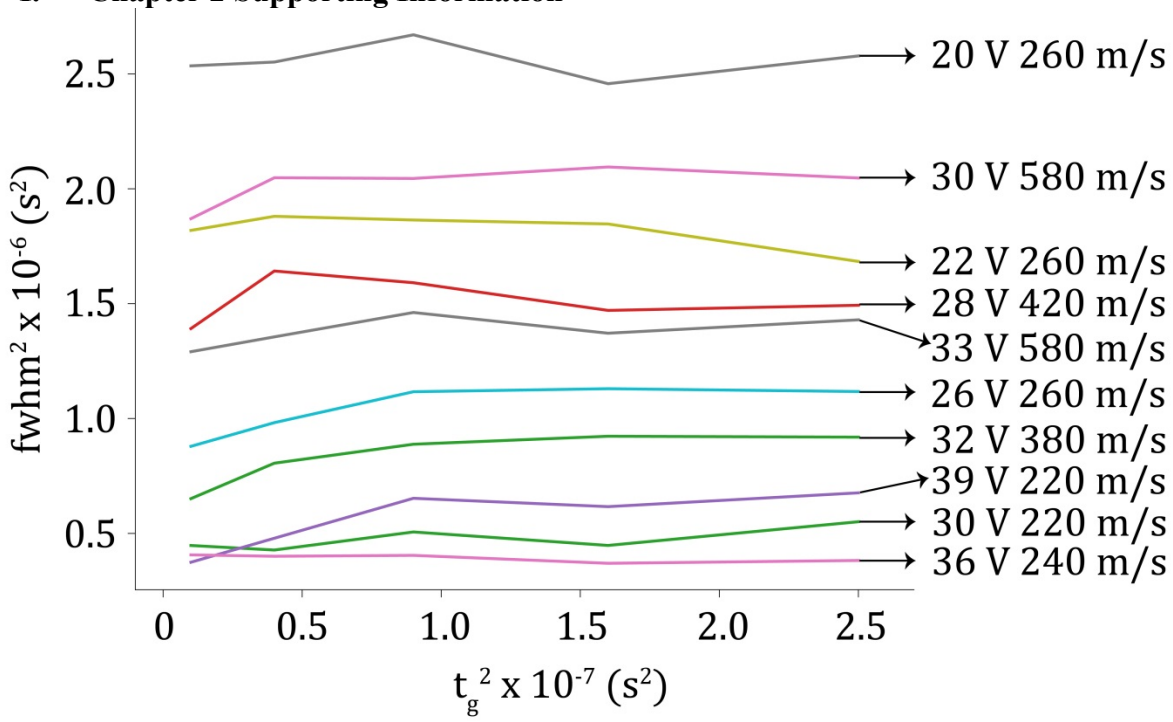


Figure 6-1. Two conformations of the same protein. Conformation shown on the left is more elongated than on the right, which yields smaller CCS values.

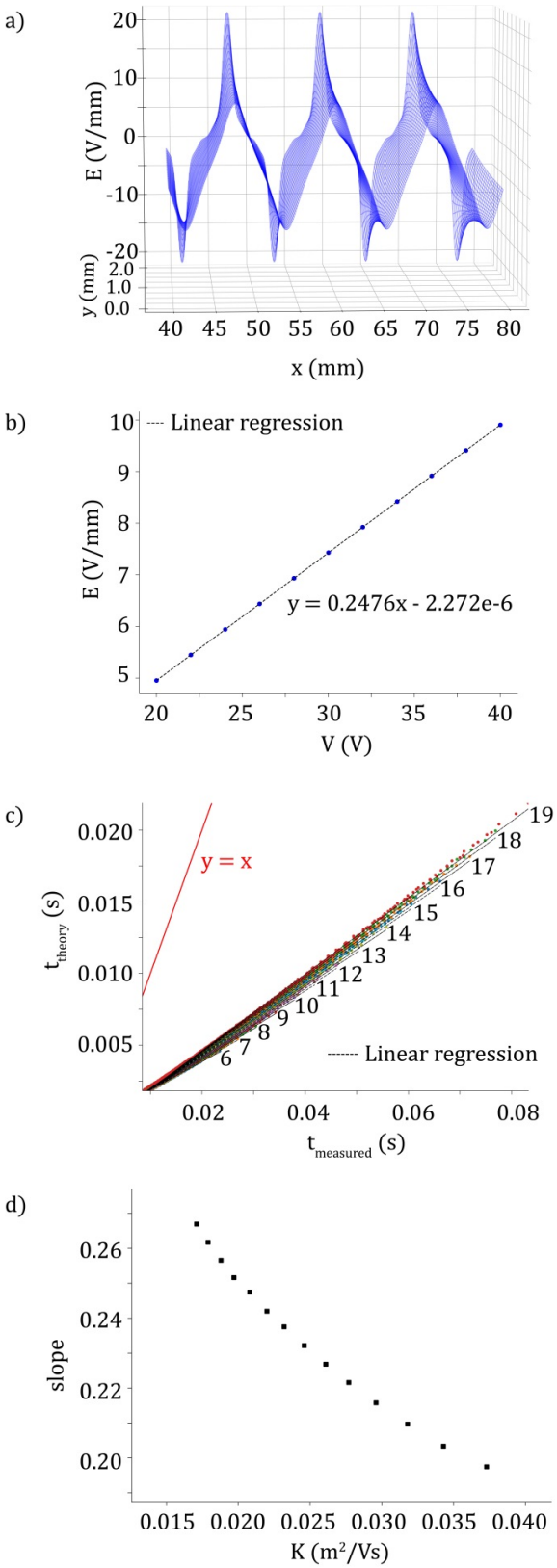
models to deliver a high accuracy method. These machine learning models have the potential to decrease the computational time and increase the accuracy for CCS calculations compared to the IMPACT method. Such a tool would enable high-throughput structural characterization and has the possibility of even on-the-fly CCS calculations during MD simulations.

Appendices

I. Chapter 2 Supporting Information



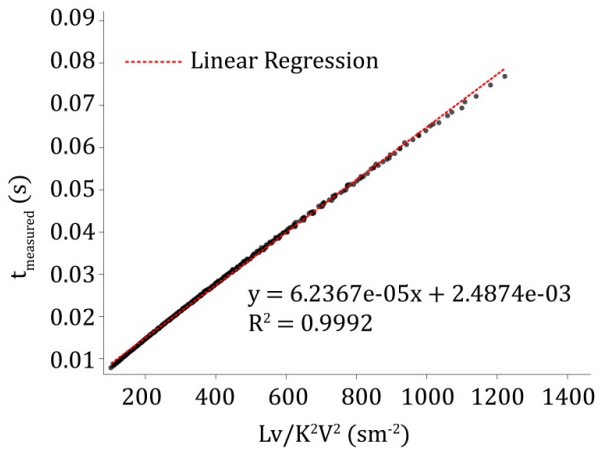
Appendix: Figure I-1. Plot of w^2 and t_g^2 for Ac-Ala_{18}-Lys^+ species at V (V) and v (m/s) conditions labeled in the plot.



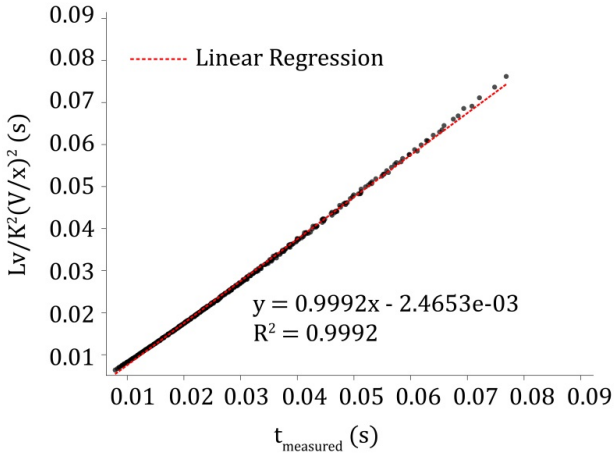
Appendix: Figure I-2. Analysis of TWIM electric field and comparison of $t_{measured}$ and t_{theory} from Shvartsburg's relationship. **a)** Plot of electric field (**E**) along traverse (**x**) and radial (**y**) direction along the TWIM cell built in SIMION 8.1 at **V** of 40 V. Linear

regression of maximum **E** in **x** direction at **y** = 0 and applied traveling wave dc voltage is shown in **b)**. This yields a function to approximate maximum **E** at any given **V**. **c)** shows the plot of measured centroid arrival time ($\mathbf{t_{measured}}$) vs. arrival time calculated using Shvartsburg's equation[236] ($\mathbf{t_{theory}}$) for the Ac-Ala_n-Lys^{1+} peptide species, with n = 6 – 19, at a range of **V** and **v**. y = x line is drawn to show the disagreement between theory and measured values. Slope from the linear regression between $\mathbf{t_{measured}}$ and $\mathbf{t_{theory}}$ is plotted as a function of ion mobility (**K**) in **d)**. This shows the difference between measured and calculated arrival times as a measure of $\mathbf{t_{theory}}/\mathbf{t_{measured}}$. This difference appears to have some correlation with **K**. We are investigating this behavior with broader range of ions and TWIM conditions, along with SIMION simulations and theoretical studies, to establish better TWIM principles.

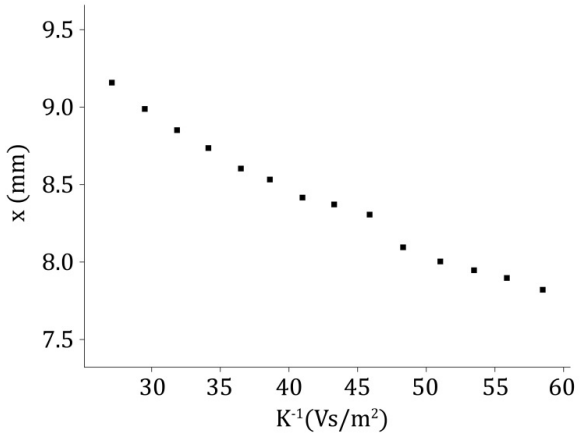
a)



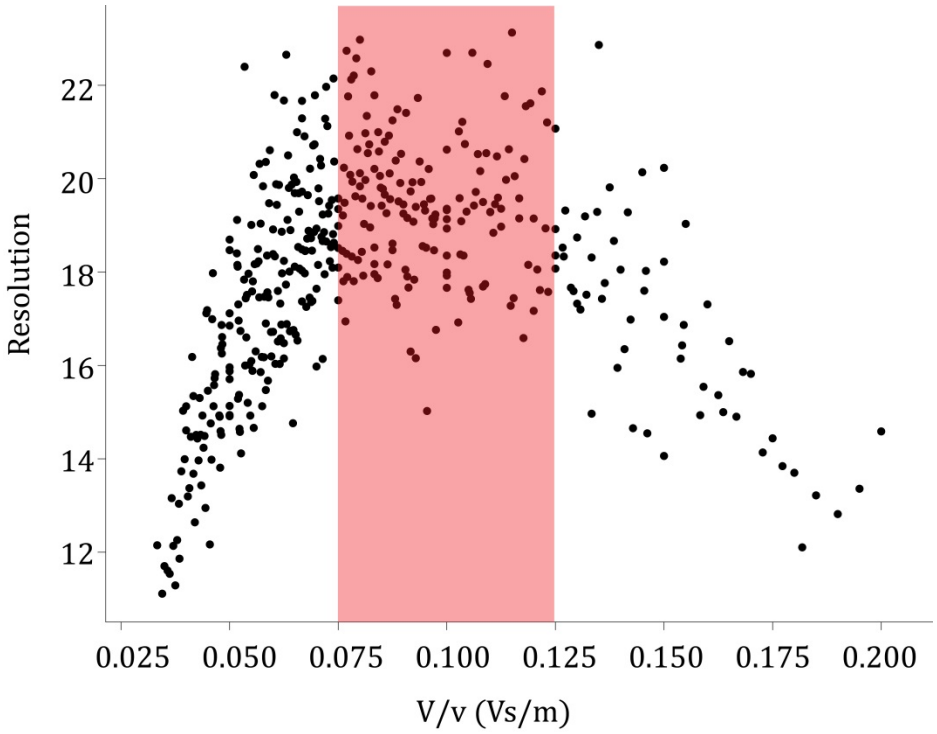
b)



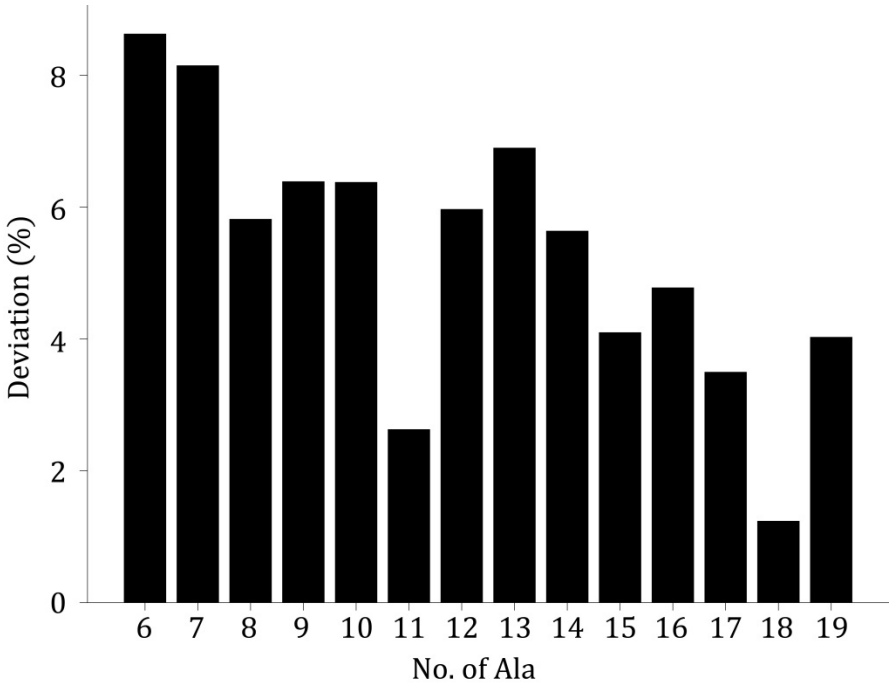
c)



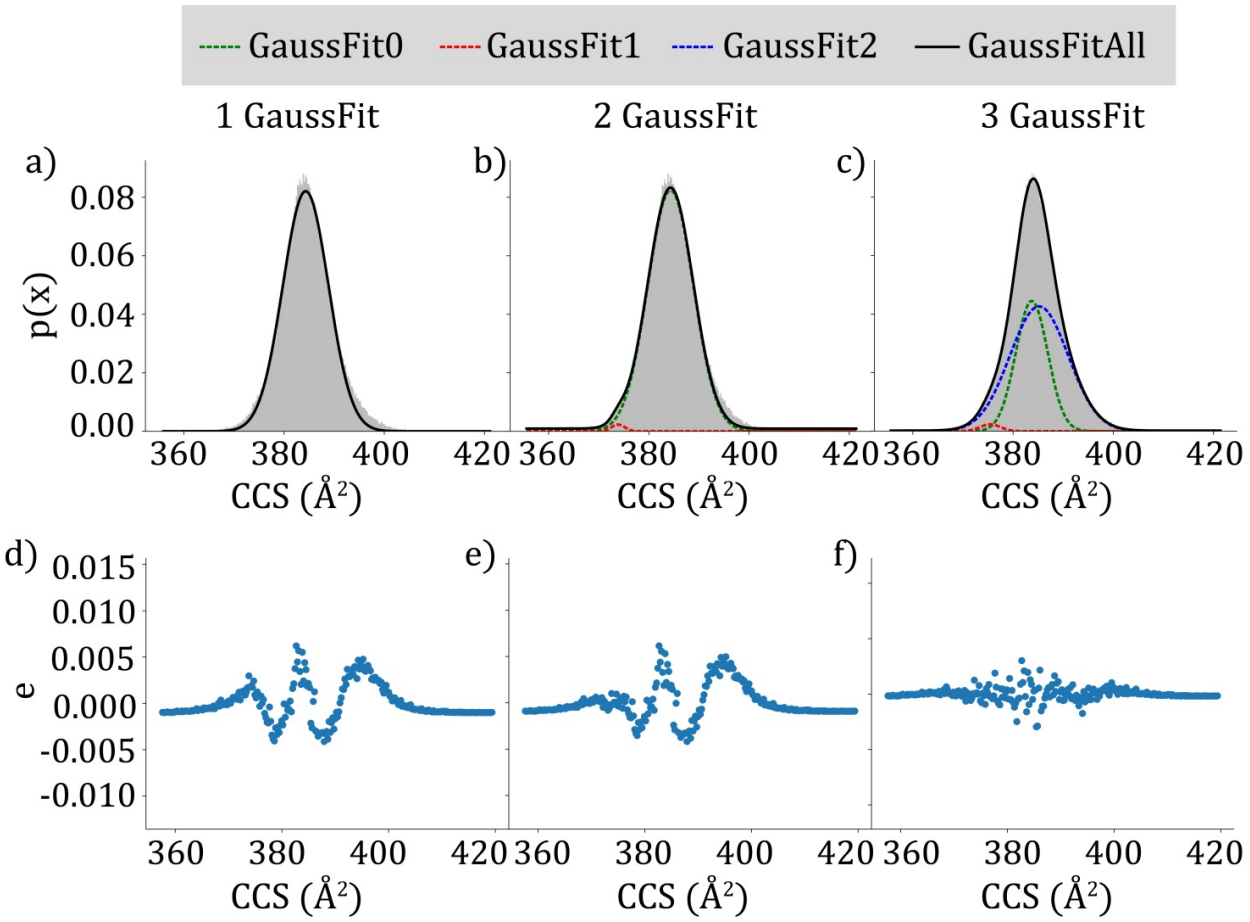
Appendix: Figure I-3. Calibration to relate $t_{measure}$ to the TWIM effective electric field. Linear regression between $\mathbf{t_{measured}}$ and $\mathbf{Lv/K^2V^2}$ is shown in plot **a)**. This linear regression model can be used to estimate the arrival time by calculating an effective **E**, which is **V/x** where **x** is the slope of the linear regression model, an ion experiences during its flight in TWIM cell. The corrected $\mathbf{t_{theory}} = \mathbf{Lv/K^2(V/x)^2}$ is plotted against $\mathbf{t_{measured}}$ and shows strong correlation, with slope of 0.99, as shown in **b)**. In **c)**, **x** is plotted against inverse mobility ($\mathbf{K^{-1}}$) that shows the relationship between **x** and the ion. We are investigating this further to better understand TWIM fundamentals.



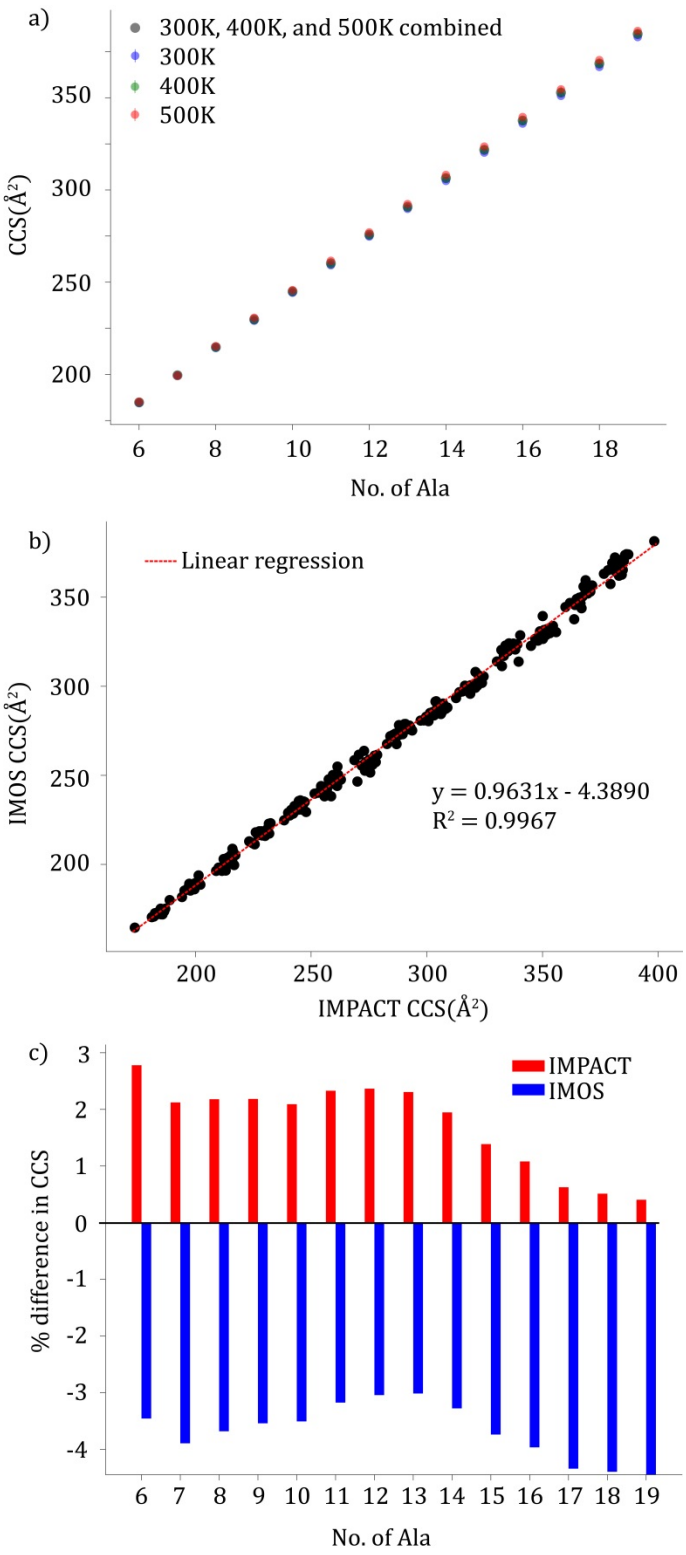
Appendix: Figure I-4. Plot of resolution as a function of **V/v** for Ac-Ala_{18}-Lys^{1+} species. The red shaded area indicates the region where maximum resolution is observed.



Appendix: Figure I-5. % deviation between measured and estimated **fwhm** is shown as a function of number of alanine in Ac-Ala_n-Lys peptides. % deviation is calculated as (1-slope) x 100 where slope is obtained from the linear regression between measured and calculated **fwhm**.

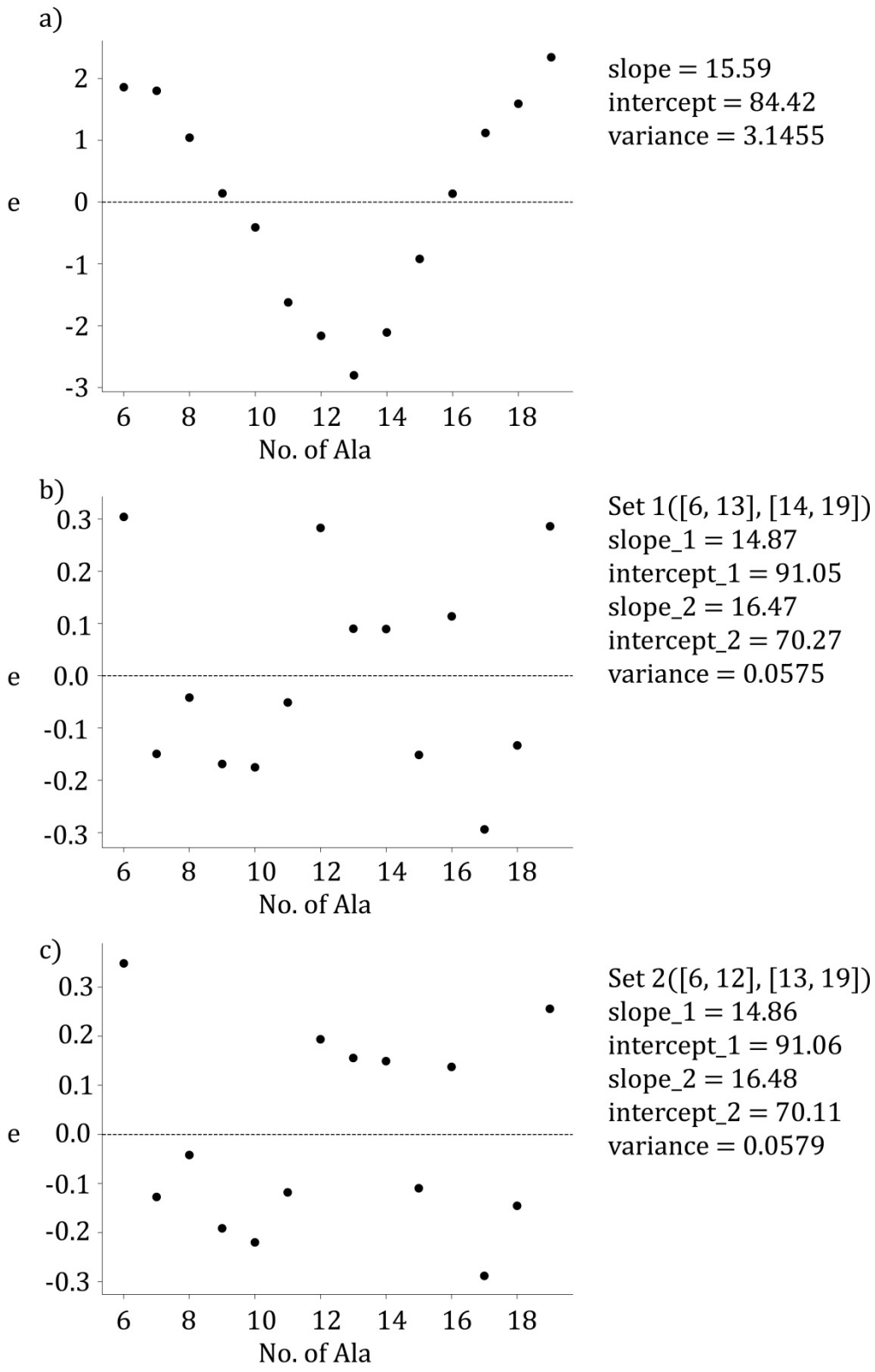


Appendix: Figure I-6. Gaussian mixture model to fit CCS distributions. **a)**, **b)**, and **c)** shows Gaussian fitting to CCS distribution with one, two and three Gaussian functions, respectively with CCS on x-axis and probability density (p(x)) on y-axis for Ac-Ala_{19}-Lys^{1+} models combined from 300 K, 400 K, and 500 K across all the replicates. **d)**, **e)**, and **f)** shows the residual (**e**) plots for **a)**, **b)**, and **c)**, respectively. It is clear that neither one nor two Gaussian functions model the distribution properly, with variance of 3.64×10^{-6} and 3.51×10^{-6}, respectively, whereas three Gaussian mixture models the distribution with lower variance of 3.74×10^{-7}. Similar fitting was performed on all the peptides for all different temperature MD runs and weighted mean and variance was calculated from the mixture model.



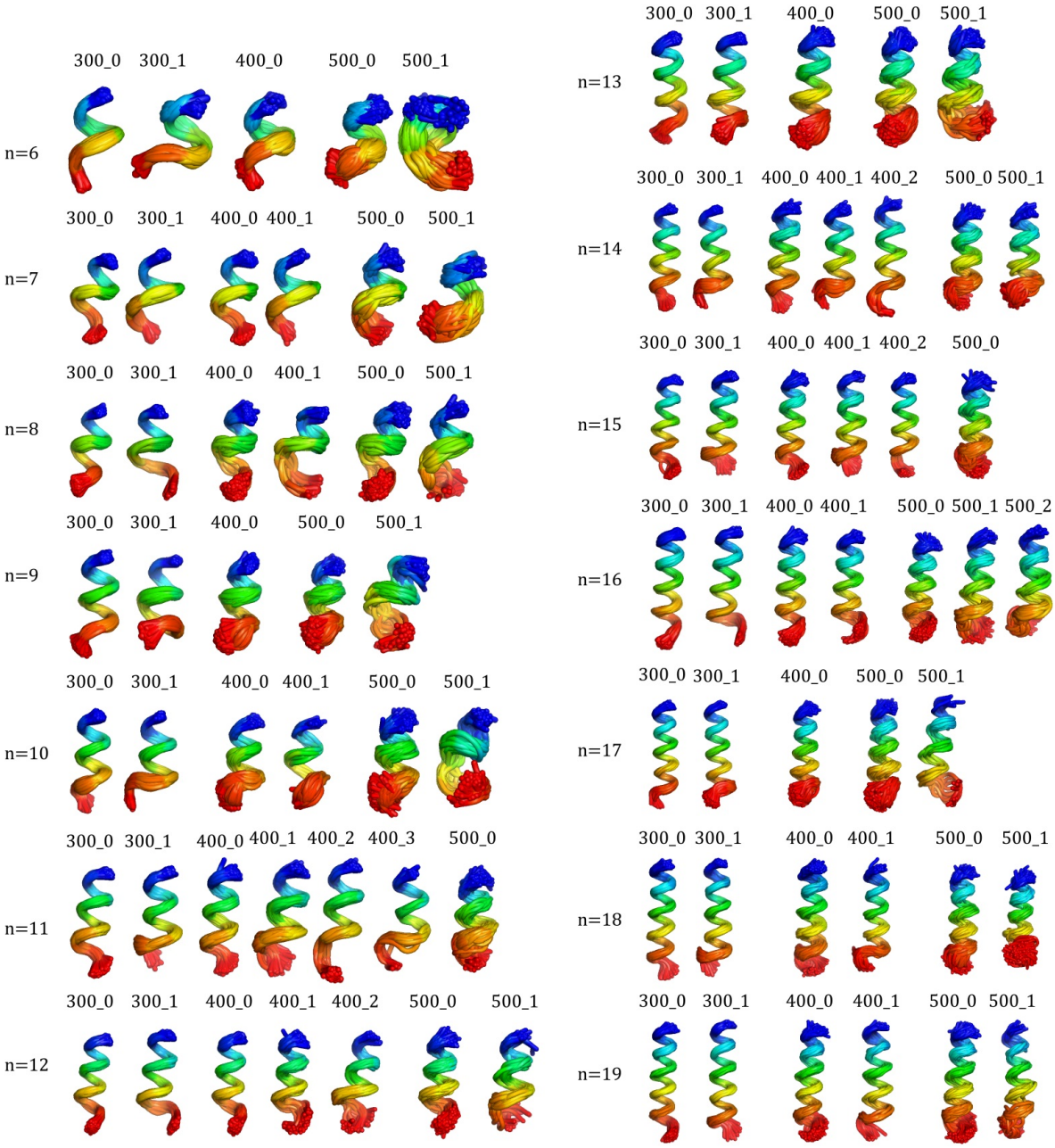
Appendix: Figure I-7. Comparison of experimental and theoretical CCS values calculated using IMPACT and IMoS. **a)** plots the weighted mean of the CCS distribution for Ac-Ala$_n$-Lys^{1+} peptide structures from 300 K, 400 K, and 500 K individual MD output as well as models combined from 300 K, 400 K, and 500 K, across all the replicates. The mean CCS do not differ significantly across different temperature. **b)** shows the the linear regression between CCS calculated using IMPACT trajectory method and IMOS trajectory method for 280 models comprising of 20 models from each Ac-Ala$_n$-Lys peptides with n = 6 – 19. The regression

model function was used to convert the weighted mean CCS from IMPACT to IMOS CCS values. **c)** shows the bar chart of % difference between experiment and IMPACT and IMOS calculated CCS. IMPACT CCS values are within 3% of experiment values. IMOS CCS values are within 5% of experiment values. Also IMOS CCS values are all underestimated due to the negative difference observed. This is due to the linear regression model generated in **b)** as it shows IMOS CCS values are consistently lower than IMPACT CCS values.

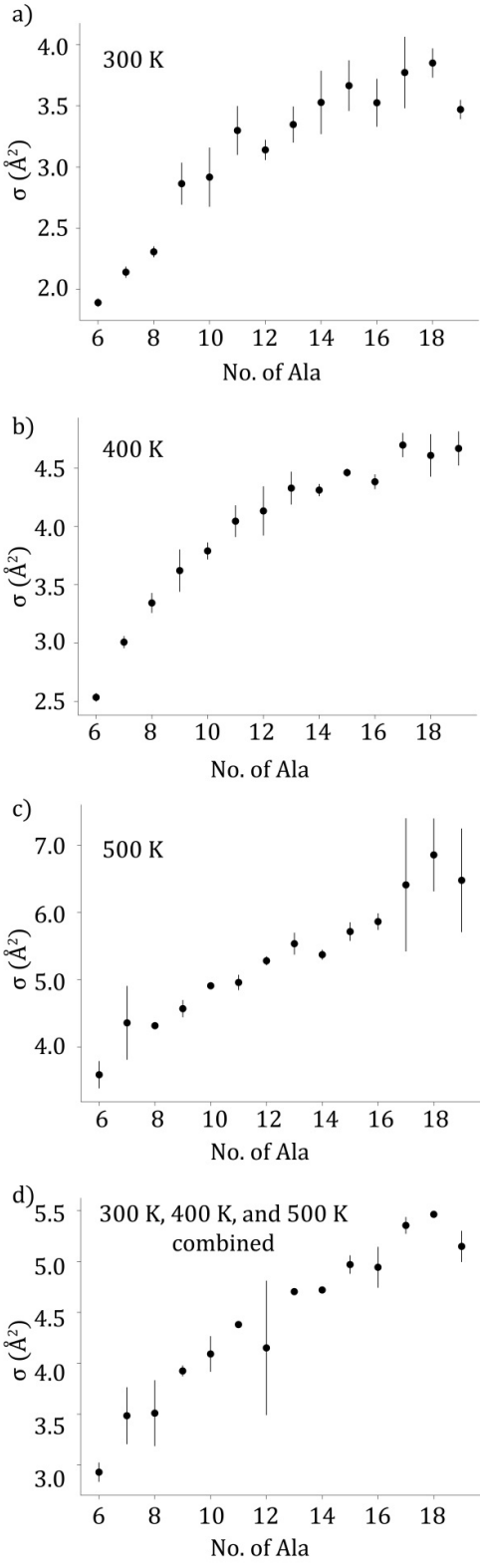


Appendix: Figure I-8. Residual analysis of a single linear regression model for the experimental CCS values against number of alanine in Ac-Alan-Lys peptides, **a)** Residual (**e**) plot when fitting a single linear regression model for the experimental CCS values against number of alanine in Ac-Ala$_n$-Lys peptides. Due to this decreasing and increasing trend of **e**, it can be inferred that more than one function is required to model the trend. **b)** Residual (**e**) plot when fitting linear regression model to two sets comprising

of n = 6 – 13 and n = 14 – 19 in Ac-Ala_n-Lys peptide. Similar analysis is done in plot **c)** with sets comprising of n = 6 – 12 and n = 13 – 19 in Ac-Ala_n-Lys peptide. For the last two cases, variance has decreased considerably when compared to fitting a single linear regression model as shown in **a)**. **b)** and **c)** both seem reasonable models as they yield similar slope, intercept, and variance values.



Appendix: Figure I-9. Cluster ensembles of Ac-Ala$_n$-Lys peptides with n = 6-19 for 300 K, 400 K, and 500 K MD simulation. Each ensemble is labeled as temperature(K)_cluster nomenclature above each ensemble figure.



Appendix: Figure I-10. Plots of weighted standard deviation (σ) of Gaussian mixture models for Ac-Ala$_n$-Lys peptides. Plots of weighted standard deviation (σ) of Gaussian mixture models for Ac-Ala$_n$-Lys peptides as a function of number of alanine for structures from **a)** 300 K, **b)** 400 K, **c)** 500 K, and **d)** combined structures from 300 K, 400 K, and 500 K.

n	K (m^2/Vs)	slope	intercept x 10^{-4} (s)	R^2
6	0.0369	0.1975	-1.0384	0.9986
7	0.0339	0.2034	-1.3051	0.9988
8	0.0314	0.2097	-1.6326	0.9990
9	0.0293	0.2158	-1.9935	0.9991
10	0.0274	0.2216	-2.3987	0.9991
11	0.0259	0.2268	-2.8686	0.9990
12	0.0244	0.2322	-3.4163	0.9989
13	0.0231	0.2375	-3.9574	0.9989
14	0.0218	0.2420	-4.2965	0.9991
15	0.0207	0.2475	-4.7424	0.9992
16	0.0196	0.2516	-5.2373	0.9992
17	0.0187	0.2566	-5.8811	0.9992
18	0.0179	0.2617	-6.6151	0.9992
19	0.0171	0.2669	-7.3535	0.9992

Table A I-1. Results from linear regression between $\mathbf{t_{theory}}$ and $\mathbf{t_{measured}}$ for Ac-Ala_n-Lys^{1+} peptides with n = 6 – 19.

n	K (m^2/Vs)	slope x 10^{-5} (m^2)	intercept x 10^{-3} (s)	R^2
6	0.0369	8.3872	0.4273	0.9982
7	0.0339	8.7087	0.5432	0.9987
8	0.0314	7.8347	0.6566	0.9990
9	0.0293	7.6314	0.7821	0.9991
10	0.0274	7.4021	0.9261	0.9991
11	0.0259	7.2807	1.0984	0.9991
12	0.0244	7.0831	1.2803	0.9991
13	0.0231	7.0087	1.3786	0.9992
14	0.0218	6.8993	1.4453	0.9993
15	0.0207	6.5533	1.8766	0.9992
16	0.0196	6.4060	2.0417	0.9992
17	0.0187	6.3157	2.2523	0.9992
18	0.0179	6.2367	2.4874	0.9992
19	0.0171	6.1171	2.7147	0.9992

Table A I-2. Results from linear regression between $\mathbf{t_{measured}}$ and $\mathbf{Lv/K^2V^2}$ for Ac-Ala_n-Lys^{1+} peptides with n = 6 – 19.

n	K (m^2/Vs)	slope	intercept x 10^{-3} (s)	R^2
6	0.0369	0.9982	-0.41400	0.9982
7	0.0339	0.9987	-0.53211	0.9987
8	0.0314	0.9990	-0.64705	0.9990
9	0.0293	0.9991	-0.77238	0.9991
10	0.0274	0.9991	-0.91552	0.9991
11	0.0259	0.9991	-1.0865	0.9991
12	0.0244	0.9991	-1.2669	0.9991
13	0.0231	0.9992	-1.3659	0.9992
14	0.0218	0.9993	-1.4329	0.9993
15	0.0207	0.9992	-1.8593	0.9992
16	0.0196	0.9992	-2.0226	0.9992
17	0.0187	0.9992	-2.2317	0.9992
18	0.0179	0.9992	-2.4653	0.9992
19	0.0171	0.9992	-2.6911	0.9992

Table A I-3. Results from linear regression between $\mathbf{t_{measured}}$ and $\mathbf{Lv/K^2(V/x)^2}$ for Ac-Ala_n-Lys1+ peptides with n = 6 – 19.

n	K (m^2/Vs)	α (V/m)	γ x 10^{-7} (s^2)	E^2/v (V^2s/m^3)
6	0.0369	1.0627	0.5908	18843
7	0.0339	1.0571	0.6548	21111
8	0.0314	1.1004	0.6767	24104
9	0.0293	1.0883	0.8129	25687
10	0.0274	1.0691	0.9125	27674
11	0.0259	1.0567	1.0302	29281
12	0.0244	1.2250	1.1726	32756
13	0.0231	1.5360	1.2903	37650
14	0.0218	1.4753	1.4716	39376
15	0.0207	1.5095	2.3382	37633
16	0.0196	1.3951	3.1064	36763
17	0.0187	1.3941	3.0581	39944
18	0.0179	1.4014	2.4884	45887
19	0.0171	1.3160	4.1011	41462

Table A I-4. This table lists the fitting parameters **α** and **γ** from fitting the empirical width equation as well as optimal $\mathbf{E^2/v}$ value for Ac-Ala_n-Lys^{1+} ions with n = 6 – 19.

n	K (m^2/Vs)	slope	intercept x 10^{-5} (s)	R^2
6	0.0369	0.9137	3.9790	0.9624
7	0.0339	0.9185	3.7546	0.9727
8	0.0314	0.9418	2.4264	0.9811
9	0.0293	0.9361	2.6604	0.9825
10	0.0274	0.9362	2.5181	0.9854
11	0.0259	0.9373	2.2202	0.9874
12	0.0244	0.9403	1.6215	0.9897
13	0.0231	0.9310	2.5802	0.9885
14	0.0218	0.9436	1.4064	0.9919
15	0.0207	0.9590	-1.5392	0.9792
16	0.0196	0.9522	-0.1177	0.9795
17	0.0187	0.9650	-3.4078	0.9825
18	0.0179	0.9876	-9.9580	0.9838
19	0.0171	0.9597	-2.9114	0.9901

Table A I-5. Results from linear regression between measured and estimated **fwhm** for Ac-Ala_n-Lys^{1+} peptides with n = 6 – 19.

	Cluster					
	0			1		
n	%	μ	σ^2	%	μ	σ^2
6	97.4	184.57	3.6547	2.6	180.45	3.9842
7	92.6	199.96	3.4543	7.4	194.65	4.9450
8	93.0	214.84	3.9013	7.0	209.30	3.5891
9	84.2	230.07	4.6423	15.8	224.18	4.0601
10	85.6	245.27	4.8144	14.4	239.58	4.7499
11	77.5	260.28	5.3715	22.5	253.47	6.0971
12	86.7	275.64	5.3422	13.3	269.09	6.8974
13	75.5	290.92	6.4183	24.5	284.93	6.2852
14	69.6	306.00	6.6743	30.4	299.97	6.4620
15	74.7	321.56	7.1503	25.3	314.92	7.5430
16	90.2	337.02	7.2304	9.8	330.91	8.0946
17	84.3	352.63	7.6124	15.7	346.33	9.0514
18	82.4	368.11	7.8646	17.6	361.70	7.4027
19	89.7	383.42	9.4495	10.3	377.18	8.2093

Table A I-6. This table includes the percentage (%) of structures, mean (μ), and variance (σ^2) of each cluster for structures from 300 K MD simulation.

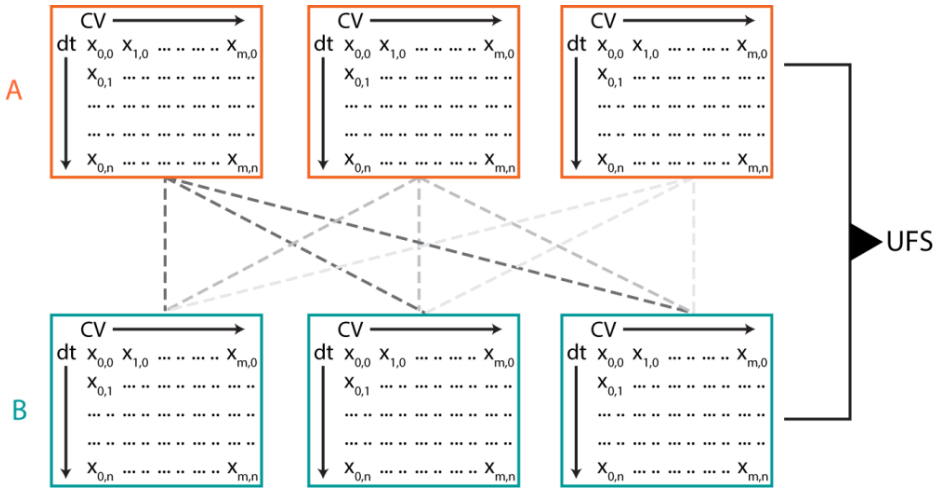
	Cluster														
	0			1			2			3			4		
n	%	μ	σ^2	%	μ	σ^2	%	μ	σ^2	%	μ	σ^2	%	μ	σ^2
6	99.3	184.74	7.4689	0.7	183.42	16.698	-	-	-	-	-	-	-	-	-
7	80.8	200.38	6.8142	17.3	195.41	8.0165	1.9	202.20	5.8518	-	-	-	-	-	-
8	97.0	214.67	14.35	3.0	217.36	34.447	-	-	-	-	-	-	-	-	-
9	99.1	229.52	14.479	0.9	224.29	5.4319	-	-	-	-	-	-	-	-	-
10	97.5	244.86	15.736	2.5	240.22	9.6941	-	-	-	-	-	-	-	-	-
11	60.0	261.42	8.8504	23.2	255.44	11.206	13.8	260.33	14.315	2.9	254.97	13.316	0.1	260.68	-
12	55.0	276.81	8.4657	37.0	275.04	23.279	8.0	270.99	17.069	-	-	-	-	-	-
13	98.1	290.90	17.856	1.9	285.90	18.122	-	-	-	-	-	-	-	-	-
14	67.1	307.96	12.339	26.0	301.50	12.035	6.1	309.66	12.275	0.7	299.67	16.109	0.1	306.05	-
15	55.8	323.75	11.642	32.0	317.25	11.72	12.2	324.85	15.586	-	-	-	-	-	-
16	74.7	339.15	11.957	24.4	333.29	17.451	0.7	330.53	22.883	0.2	343.80	6.3991	-	-	-
17	99.9	352.47	23.836	0.1	340.79	-	-	-	-	-	-	-	-	-	-
18	92.9	368.58	20.794	7.1	363.99	30.633	-	-	-	-	-	-	-	-	-
19	97.4	384.53	23.629	2.6	378.44	21.885	-	-	-	-	-	-	-	-	-

Table A I-7. This table includes the percentage (%) of structures, mean (μ), and variance (σ^2) of each cluster for structures from 400 K MD simulation. Clusters with less than 2% structures are shaded in grey.

	Cluster											
	0			1			2			3		
n	%	μ	σ^2	%	μ	σ^2	%	μ	σ^2	%	μ	σ^2
6	81.9	185.41	14.355	18.1	184.24	23.896	-	-	-	-	-	-
7	92.3	200.82	18.574	7.6	200.35	40.923	0.1	200.98	-	-	-	-
8	95.5	215.31	19.434	4.2	214.92	41.406	0.3	211.63	9.756	-	-	-
9	92.2	230.99	21.591	7.8	227.86	38.054	-	-	-	-	-	-
10	89.5	246.10	26.006	10.5	242.75	36.837	-	-	-	-	-	-
11	98.8	261.93	30.403	1.2	260.32	16.332	-	-	-	-	-	-
12	95.8	277.27	28.356	4.2	273.42	38.189	-	-	-	-	-	-
13	90.4	292.82	29.478	9.6	288.84	46.730	-	-	-	-	-	-
14	95.2	308.74	29.723	4.4	304.49	21.312	0.4	308.89	5.1987	-	-	-
15	99.9	323.76	33.256	0.1	324.67	11.72	-	-	-	-	-	-
16	83.8	340.09	31.953	11.4	336.59	51.285	4.5	333.59	55.224	0.3	344.89	53.393
17	99.0	355.96	34.049	1.0	349.67	59.996	-	-	-	-	-	-
18	75.4	371.42	34.884	24.6	365.51	70.029	-	-	-	-	-	-
19	96.9	387.14	35.192	2.8	388.12	71.194	0.3	392.06	13.169	-	-	-

Table A I-8. This table includes the percentage (%) of structures, mean (μ), and variance (σ^2) of each cluster for structures from 500 K MD simulation. Clusters with less than 2% structures are shaded in grey.

II. Chapter 3 Supporting Information



Appendix: Figure II-1. Univariate feature selection, comprised of performing F-test statistics[281] to evaluate the significance of each collision voltage in differentiating the training data classes, is performed as the first step in classification workflow. Scheme S1 shows UFS for two groups A and B, with F statistics calculated for each data set in each group, as shown by the dotted lines. F ratio is calculated as described below, and *F* ratio is then converted to p-value using the F-distribution.

Text II-1

F-test calculation

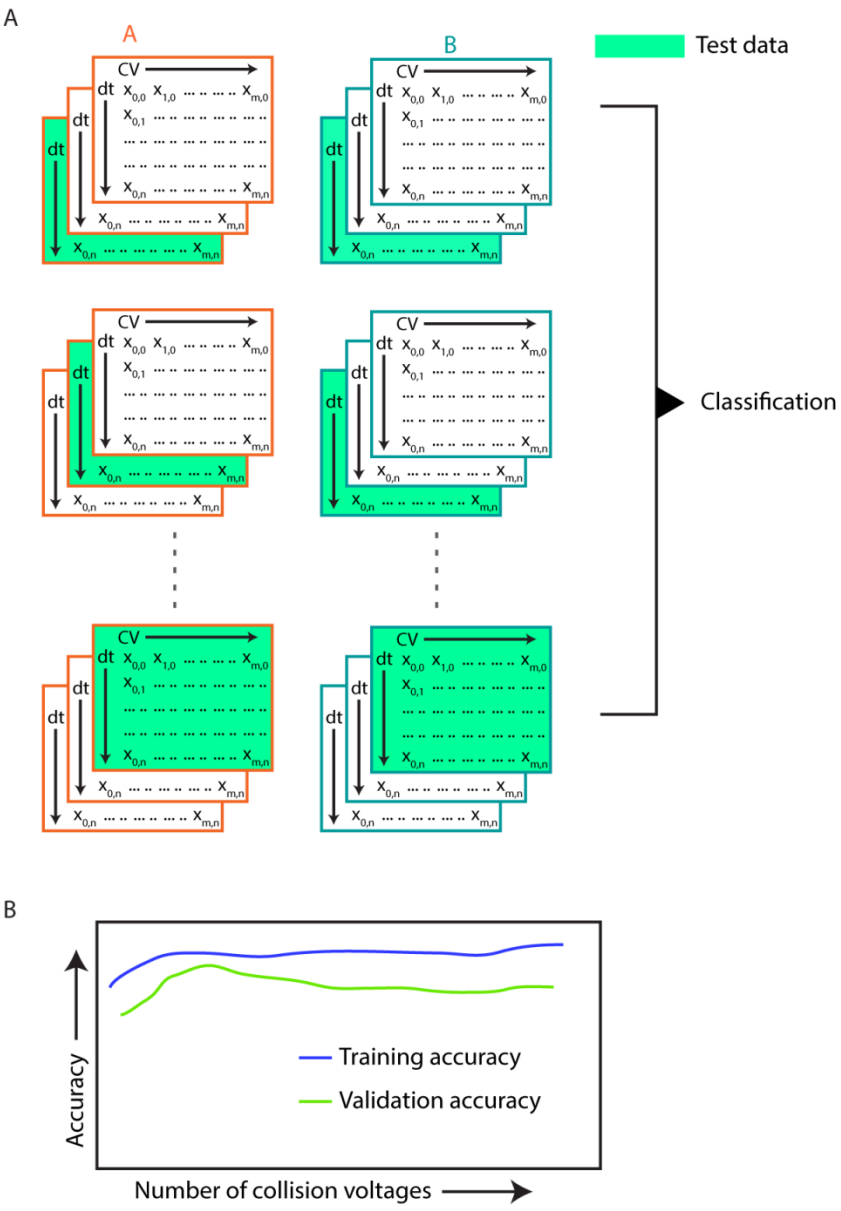
$$F = \frac{Between\ group\ variability}{Within\ group\ variability}$$

$$Between\ group\ variability = \sum_{i=1}^{K} n_i\,(\overline{Y_i} - \overline{Y})^2/(K-1)$$

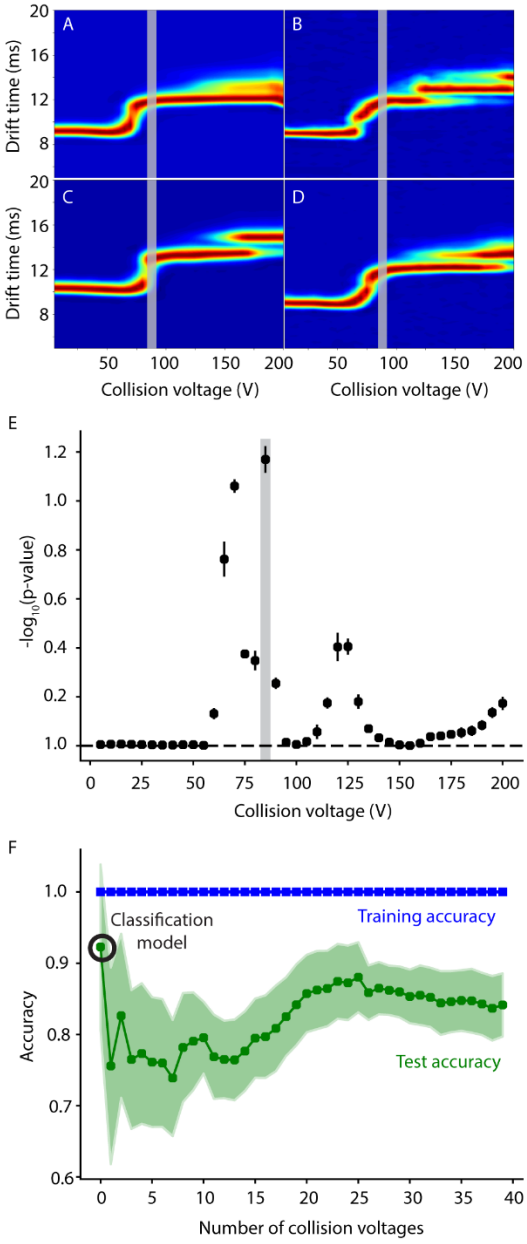
where K is the number of groups, n_i and $\overline{Y_i}$ are the number of observations and the mean of the i-th group, respectively, and $\overline{Y}$ is the overall mean of the data.

$$Within\ group\ variability = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y_i}\right)^2/(N-K)$$

where Y_{ij} is the j-th observation in the i-th group and N is the overall sample size.



Appendix: Figure II-2. Cross-validation workflow. A "leave one out" cross-validation[282] is performed by implementing the classification workflow, which consists of linear discriminant analysis (LDA)[283] followed by construction of a support vector machine (SVM)[284], across a range of sub-sections of CIU datasets corresponding to sequential addition of arrival time distributions from individual collision voltages in decreasing order of UFS score. Classification accuracy of training and test data sets are then plotted as a function of the number of collision voltages to determine the optimal model selection. A) Scheme showing "Leave one out" cross validation approach where a single CIU dataset is treated as test data in each group and the remaining datasets are used as training data. All possible combinations of training and test data set are created. The training set is used to build the classification model and the test data is used to validate the accuracy of the classification model. B) Accuracy of training and test data sets are plotted as a function of a range of CIU datasets that has sequential addition of collision voltages. The number of collision voltages resulting in greatest validation accuracy is used as the final classification scheme.



Appendix: Figure II-3. Feature selection and cross-validation for IgG1, IgG2, IgG3, and IgG4 classification from primary text Figure 3. A, B, C, and D are representative CIU fingerprints of IgG1, IgG2, IgG3 and IgG4, respectively. E) UFS results. Plot of $-\log_{10}$(p-value) against collision voltage assessing the significance of each collision voltage in differentiating the groups. 85 V, which has the highest score, is highlighted in grey and is also show in CIU fingerprints (A, B, C, and D). F) Cross-validation plot showing classification accuracy for training (blue) and test (green) data set with accuracy on y-axis and number of collision voltages on x-axis. The shaded region indicates the standard deviation from all the combinations of training and test datasets created. In this example, a single collision voltage (85V, circle marker) had the highest validation accuracy and was chosen as the final classification scheme to generate the data in primary text Figure 3.

IgG Subclass	Replicate	Probability for IgG1	Probability for IgG2	Probability for IgG3	Probability for IgG4
IgG1	1	0.69	0.12	0.06	0.13
	2	0.69	0.12	0.06	0.13
IgG2	1	0.33	0.52	0.06	0.09
	2	0.31	0.55	0.06	0.09
IgG3	1	0.07	0.04	0.73	0.16
	2	0.07	0.04	0.73	0.16
IgG4	1	0.14	0.07	0.10	0.69
	2	0.15	0.07	0.10	0.69

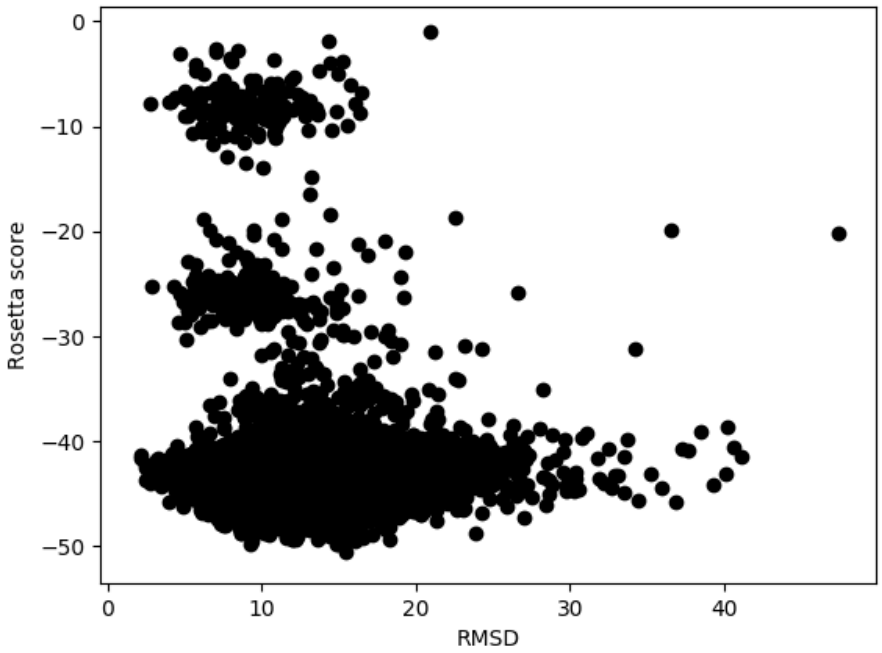
Table A II-1. Test data probability values for classification of IgG subclasses. Each subclass has higher probability values when it is classified into its respective type, as highlighted in green.

III. Studying the interactions between SERF and amyloid-β using docking.

Recently, a small 68 amino acid intrinsically disordered peptide, named MOAG-4, was discovered as an in vivo enhancer of amyloid formation in *Caenorhabditis elegans*.[291] The deletion of MOAG-4 leads to fewer amyloid formation in a *C. elegans* model of Huntington's disease. The human orthologue, small ERFK-rich factor 1 (SERF1a), have been shown to accelerate amyloid formation with a broad range of amyloidogenic proteins in vitro.[292] However, the molecular mechanism by which SERF promotes aggregations is still unclear. It is hypothesized that interaction of amyloidogenic protein with SERF acts as a catalyst providing the ideal interface for amyloid growth.[292] Therefore, we investigated the interactions between SERF and amyloid-β to help understand the aggregation boost.

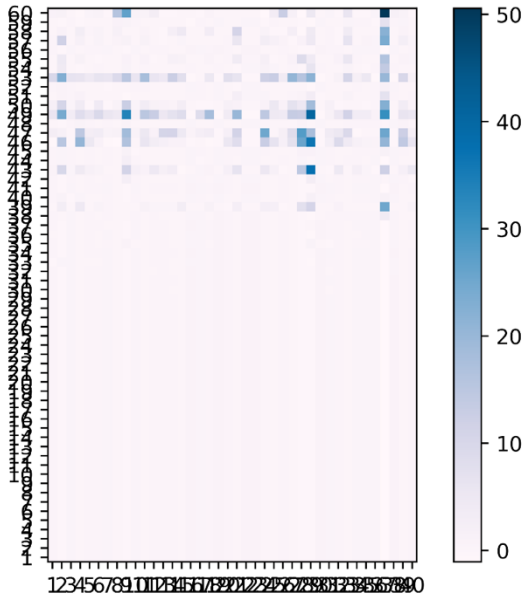
Appendix: Figure III-1. Yeast SERF models created using homology modeling.

We created a yeast homolog of SERF using SWISS-MODEL. Six different structures were scored the best (Figure III-1). We proceeded with modeling of the interaction of amyloid-β with all the structures of yeast SERF using Rosetta Docking protocol.[293] The docking was performed to create 10,000 decoy structures. (Figure III-2).



Appendix: Figure III-2. Plot of Rosetta score vs. RMSD of 10,000 structural decoys from docking.

Because convergence of structures was not apparent (Figure III-2), we clustered a randomly selected 20,000 structures using root mean square deviation (RMSD) as the distance matrix. We then examined interaction patterns using contact analysis. Contact maps were created using a 5Å distance cut-off between alpha carbons of the interacting partners. Within each cluster, the contact maps were summed to calculate the frequency of residues that participated in the interaction. We then created a score for a pair of interacting residues using the following:

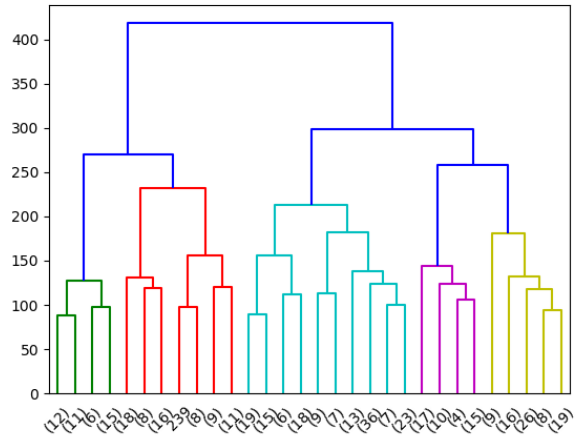


Appendix: Figure III-3. A 3D heat map showing the interaction score using equation 1 with amyloid-β residues on the x-axis and SERF residues on the y-axis.

$$score_{ij} = w_{ij} * (w_i + w_j) \qquad III - 1$$

where i and j are the residues of amyloid-β and SERF, respectively, and w is the z-score. An example of such score is shown in Figure III-3.

We used the interaction score defined above to highlight the location preference for binding. In each cluster, we selected the top 10 best scoring interaction pairs. Since there were 6 yeast SERF clusters and each cluster had an average of 5 amyloid-β and SERF interaction clusters, a total of 392 pairs of interaction were selected. We then performed hierarchical clustering on the interacting pair of residues using their sequence number to classify different interactions (Figure III-4).



Appendix: Figure III-4. Cluster of interaction pairs using their binding location.

The clusters are separated according to a wide locality of binding regions. A brief summary of them are listed below:

1. C-terminal residues of amyloid-β interacting with c-terminal part of SERF
2. N-terminal region of amyloid-β interacting with mid region and c-terminal of SERF
3. Few interactions where both SERF and amyloid-β interacts using their n-terminal region

SERF has a more preferred location of binding, either at the c-terminal or the mid region, than amyloid-β. This suggests that SERF may act as a chaperone for promoting oligomer formation in amyloid-β.

With such plenty of data from our modeling efforts, we plan to examine the interactions in more detail (polar, hydrophobic, electrostatic, etc.). A handful of top scoring interactions will be used for alanine scanning to disrupt the interactions and examine the formation of SERF and amyloid-β complexes using IM-MS. Such experiments will help identify precise locations of

SERF and amyloid-β interactions along with providing useful insights into the mechanism of aggregation promotion by SERF.

IV. Ion mobility mass spectrometry reveals the structures and binding modes of islet amyloid polypeptide with lipid bilayers

Introduction

The formation of insoluble deposits from soluble peptides and proteins has been an area of significant scientific interest in disciplines ranging from physics to biology and medicine. This is primarily due to the association of many neurodegenerative disorders and diseases with such deposits.[294] Alzheimer's disease, which is a neurodegenerative disease associated with the aggregates of Amyloid-β (Aβ) peptide, is a highly prevalent cause of dementia with an estimated 5.7 million Americans suffering from the disease in 2018.[295] A recent review highlighted that 37 peptides or proteins have been found to form amyloid aggregates in human pathologies.[294] However, the biophysical forces that drive the formation of such aggregates from soluble proteins are still unclear. Protein molecules adopt a multitude of different conformational states whose thermodynamics and kinetics are determined by the protein sequence as well as interactions with other biomolecules in a highly complex cellular environment.[296] Evolution has optimized protein structures and dynamics to conserve their functionality.[297] However, it is evident that there are circumstances where proteins have evolved to form dysfunctional aggregates.[298] The early stages of aggregation are typically marked by unstable assemblies that have clusters consisting of dynamic monomers.[299] Therefore, depending on the structural dynamics of available monomers, aggregates may adopt highly disordered, partially folded, or native-like ordered oligomeric forms. As aggregation proceeds, further internal reorganization can lead to the formation of stable β-sheet structured oligomers. These β-structured oligomers can grow by self-association and form well

defined highly ordered amyloid fibrils.[300,301] In contrast, the disordered or native-like aggregates may not undergo any structural reorganization and grow to form large amorphous deposits or native-like assemblies.[302] These large aggregates and fibrils are associated with human disease as they accumulate in various neuropathic, hereditary, and systemic pathologies.[294]

Disease symptoms for many neurodegenerative diseases are thought to arise from abnormal cell death where the amyloid deposits are localized. Although autophagy has been implicated in many neurodegenerative diseases including Alzheimer's disease, Parkinson's disease, and Huntington's disease, its exact role in neuropathological conditions remains unclear.[303–305] However, the processes of fibril and/or aggregate formation are suspected to induce cytotoxicity leading to the onset of symptoms.[306] There is a common cytotoxicity mechanism for all amyloid diseases where cell homeostasis is disrupted due to damage associated with cellular membranes.[307] Two generic pathways have been suggested for how amyloidogenic peptides or proteins trigger cell apoptosis.[308] First, amyloidogenic peptides can form toxic oligomeric species and embed themselves into the cell membrane. This could lead to the formation of unregulated pores in the architecture of the cell, affecting its homeostasis. Second, the growth of oligomers and fibrils within the membrane may act to permeabilize the cell, severely affecting the its function. These general options rely on a great number of molecular-level details considered important in the action of amyloid peptides and proteins, each of which are thought to act somewhat differently in the context of their respective disease states.[308]

Here, we focus on Islet Amyloid Polypeptide (IAPP) oligomers, which are implicated in type 2 diabetes mellitus (T2DM).[309] IAPP is an amyloidogenic 37 residues long peptide hormone produced by β-cells of the islets of Langerhans in the pancreas.[308,310] There is evidence of human IAPP (hIAPP) amyloid deposits in the pancreas of T2DM patients, and elevated levels of IAPP in

T2DM has been linked to the death of insulin-producing β-cells.[311] Therefore, interactions between hIAPP and β-cell membranes are thought to play a role in β-cell degradation. Like other amyloidogenic peptides, hIAPP may induce cell apoptosis by disrupting cell homeostasis through a variety of mechanisms.[312] One general strategy for therapeutic interventions in the context of amyloidogenic peptides is to reduce the population of peptide oligomers by directly disrupting the processes that lead to their formation.[313] However, the mechanism of hIAPP membrane permeation is not clearly understood. Recently, a study by Martel et al[314] characterized the structures of IAPP aggregates and their interactions within a membrane bilayer. Their results suggested that IAPP oligomers are cytotoxic species that, without forming pores, interact peripherally on the outer leaflet of membrane bilayer to destabilize the cellular membrane. Although this study provided significant insights into potential IAPP cytotoxicity, detailed structures of IAPP oligomers were not recovered from this work, and such data would be greatly enabling for developing new therapeutic strategies associated with T2DM.

In this study, we employ native ion mobility-mass spectrometry (IM-MS) to study the interaction of IAPP with lipid nanodiscs (NDs). Rat IAPP (rIAPP) was used as our model system due to its high sequence and structure similarity to hIAPP, while exhibiting slower amyloid fibril formation kinetecs.[308,314–316 317] Once ND-rIAPP oligomers are transferred into the gas phase using nano electrospray ionization (nESI), we used collision induced dissociation (CID) to liberate various lipid-bound rIAPP oligomers from the ND, recording stoichiometries and collision cross-sections (CCSs) for all ions observed. By carefully analyzing trends in CCS growth as a function of lipid attachment, we conclude that most of the available rIAPP dimer population interacts only with the surface of the ND. We confirm this observation though the construction of all-atom models of the IAPP-ND complexes as well as through detailed comparisons with CCS data for

gramicidin A dimer bound to NDs. Gramicidin A is well-known to insert into lipid bilayers[318], and displays a contrasting CCS growth curve to our IAPP-lipid complex data, which we rationalize based on protection of the gramicidin A dimer from gas-phase structure re-arrangements within the ND. Overall, our data is some of the first to ascertain the details IAPP-ND interactions, and illustrates the potential of IM-MS data for carrying out similar experiments on a wide range of amyloidogenic peptide systems in the future.

Materials and Methods

Samples: rIAPP was purchased from Anaspec, Inc and dissolved in DMSO at a final concentration of 5mM. After 24-hour incubation at 25°C, stock solutions were diluted into 20 mM ammonium acetate, pH 6.8, to a final concentration of 400 μM. Gramicidin A dimer (Sigma-Aldrich) was dissolved in methanol at 400 u\μM concentration. 1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) was purchased from Avanti Polar Lipids. 22A peptide was synthesized using Fmoc-based chemistry and purified by HPLC. ND was prepared using the established protocol described elsewhere.[317] IAPP and gramicidin A were mixed with ND at a ratio of 20:1 (peptide:nanodisc) and incubated for 24 hours at room temperature. The mixture was then analyzed using IM-MS.

IM-MS instrumentation: A Waters Synapt G2 IM-MS instrument was used for analyzing ND-peptide complexes. Briefly, the samples were ionized using nESI with a capillary voltage of 1.1-1.3 kV, cone voltage of 50 V, temperature of 100°C. Trap collision voltage of 40 V was applied to liberate peptide species from ND. Traveling wave ion mobility was operated at a pressure of 3.4 mbar (He and N_2 flow were set at 200 and 90 ml/min, respectively). IM arrival time distribution

(ATD) data was taken at 300 m/s wave velocity and 28 V, 30 V, and 32 V wave amplitude values. Data was acquired for almost 10 minutes to get a good signal to noise ratio.

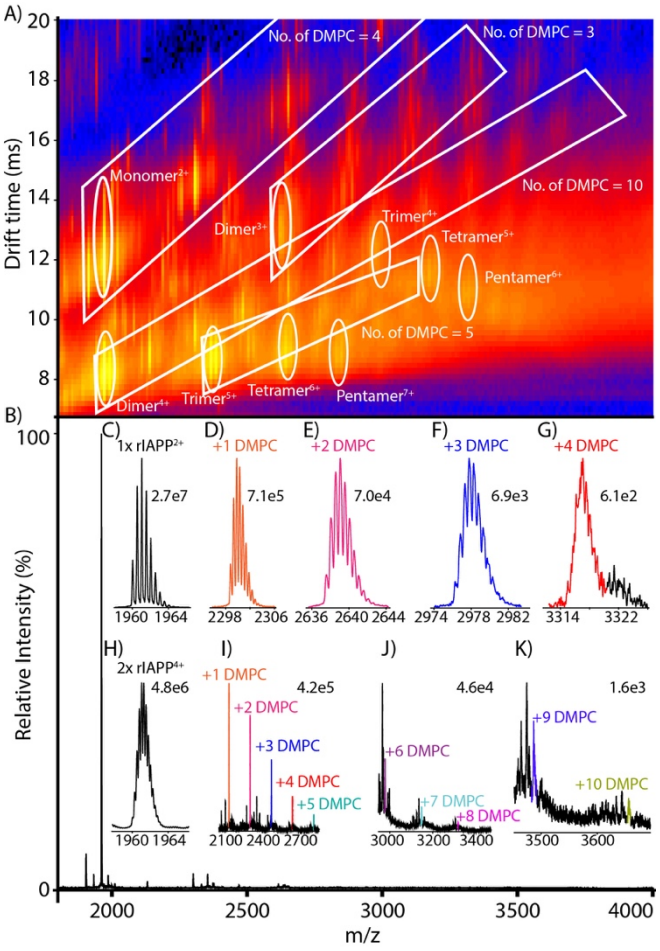
IM-MS data analysis: MassLynx (Waters) and DriftScope (Waters) were used to obtain m/z distribution list. MMass[319] was used to deconvolute mass spectra when isotopically resolved peaks were present. When m/z distribution was not isotopically resolved, MMass was used to fit the m/z peaks and obtain centroid m/z values. An in-house python script was used to annotate the peak list. Briefly, the theoretical m/z list was generated for all possible combinations of adducts with the peptide. Adducts were assigned with a mass and maximum number allowed. The theoretical m/z list was matched with the experimental m/z list with 20 ppm and 50 ppm for with and without isotopically resolved peaks, respectively. After high-confidence peak assignments were achieved, ATDs were extracted from each m/z distribution. CCSs were determined using the TWIM calibration method described in chapter 4.[244]

rIAPP modeling: rIAPP (pdb id:2KJ7) was used to create a dimer model using Rosetta docking. Briefly, 30,000 structures were generated using the docking workflow that consisted of low-resolution and high-resolution protocol with all-atom energy minimization.[320] rIAPP dimer was placed in a DMPC constituted ND in both integral and peripheral orientations using CHARMM-GUI.[321,322] Six cycles of equilibration step were performed followed by a 5 ns production run. At this point, the root mean square fluctuation of rIAPP dimer was converged. We used the rIAPP dimer structure from the ND complex to generate gas-phase models. rIAPP dimer was equilibrated in vacuum for 200 ps using CHARMM 36 force field. Following equilibration, a 20 ns constant temperature run at 300 K was performed to get the vacuum-relaxed structures. Additionally, a 20

ns replica exchange MD was performed with 12 replicas at temperatures that were exponentially distributed between 300 and 800 K. 300 K replica was used to generate *in-vacuo* structures of rIAPP dimer.[66] 2000 structures were extracted from the last 10 ns of the simulation. Theoretical CCS were calculated for those structures using IMPACT.[137]

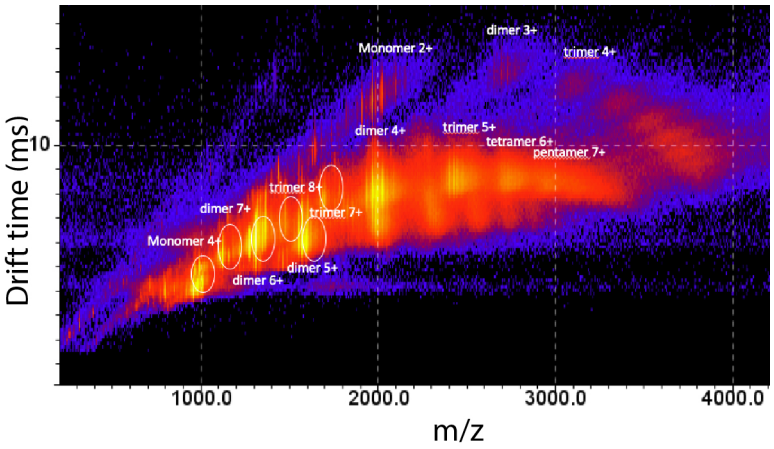
Results and Discussion

Recent work[323] has demonstrated that native MS can be used to directly detect the incorporation of antimicrobial peptides within NDs, but remained unable to directly ascertain the stoichiometry and structures of the oligomers present. In contrast, we utilize CID in tandem with IM-MS analysis to eject complexes from the ND in order to directly measure peptide oligomer mass and CCS using IM-MS. Figure IV-1A shows a plot of IM drift time versus m/z, where signals for a range of rIAPP oligomeric and lipid-bound states are detected following rIAPP-ND complex CID. IM aids in separating different oligomeric species and their lipid-bound states, enabling a more detailed interpretation of the mass spectrum acquired (Figure IV-



Appendix: Figure IV-1. rIAPP oligomeric and lipid-bound states detected after peptide-ND complex underwent CID. **A)** A 3D contour plot with m/z on the x-axis, drift time on the y-axis, and ion intensity represented by the color scheme with bright yellow indicating high intensity and blue indicating low intensity. The white circle indicates the various oligomeric states of the rIAPP molecule. Each white box contains signals for one oligomeric species and its lipid-bound states. **B)** The integrated mass spectrum combining all the signals from **A)**. It is rather challenging to identify the peaks using this mass spectrum profile. Therefore, signals were extracted from each white box. Mass spectra for monomer 2^+ ions with 0 to 4 lipid-bound states are shown in **C)** to **G)**. Similarly, **H)** through **K)** shows the mass spectra for dimer 4^+ ions with 0 to 10 lipid-bound states. Intensity values for mass spectra are listed on the top right corner for **C)** through **K)**.

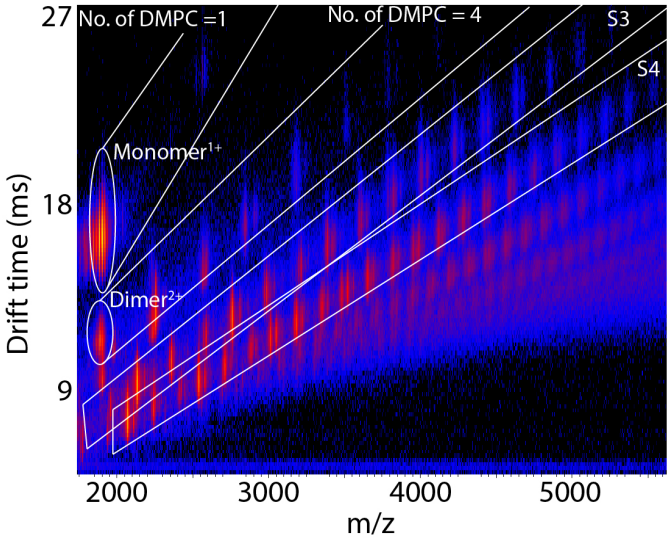
1B). For example, we detect a number of clear trend-lines in the IM-MS data collected for IAPP complexes ejected from NDs, and upon detailed investigation, we are able to assign these trends to separate groups of IAPP monomers or oligomers bound to different numbers of DMPC lipids. Specifically, our data contains signals corresponding to rIAPP monomers and oligomers up to, and including, pentamers. Similar oligomeric states can be detected for rIAPP samples prepared in the absence of NDs (Figure IV-2). In addition, Figure IV-1C – IV-1G shows detailed peak assignments for the rIAPP monomer 2+ series, where we identify signals corresponding to between 0 – 4 DMPC lipids bound. IM-MS data acquired for IAPP-lipid complexes released from NDs also reveals that rIAPP dimers bind to a greater number of lipids than rIAPP monomers. For example, we identify rIAPP dimer 4+ signals corresponding to between 0 – 10 lipid-bound species (Figure IV-1H and IV-1K). Despite the large number of signals and significant chemical noise in the dataset shown in Figure IV-1, we have developed high-confidence identifications for most of the signals detected.



Appendix: Figure IV-2. A 3D IM-MS plot showing the oligomeric states of rIAPP in absence of ND.

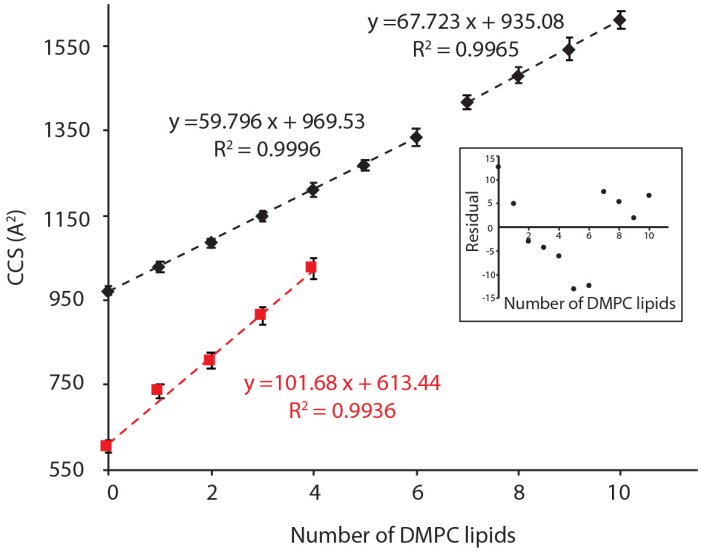
Similarly, we also identified gramicidin A oligomers and their lipid-bound states from the gramicidin A-ND complex. Figure IV-3 shows gramicidin A monomer^{1+} and dimer^{2+} species with one and four DMPC lipids bound, respectively. The selection S3 and S4 each contain a mix of oligomers and their lipid bound states. Selection S3 contains signals for 3^{+} ions representing oligomeric and lipid bound states of gramicidin A dimer, trimer, tetramer, and hexamer oligomers.

In selection S4, similar pattern of gramicidin A oligomers and lipid bound states are observed but with a charge of 4^+. Unlike rIAPP, the signals observed in Figure IV-3 contains gramicidin A with methylation and adducts such as sodium ion and ammonium ion. For this study, we decided to focus just on the protonated species to make comparisons with rIAPP easier.



Appendix: Figure IV-3. Oligomers and lipid bound species of gramicidin A after liberating from ND. A 3D contour plot with m/z on the x-axis, drift time on the y-axis, and ion intensity represented by the color scheme with bright yellow indicating high intensity and blue indicating low intensity. Monomer1^+ and Dimer2^+ species are labelled with the number of DMPC bound states. Selections S3 and S4 contain series of ions belonging to 3^+ and 4^+ charge states.

In analyzing our peptide-lipid complex datasets, we reasoned that the number of lipids attached to the peptide oligomers observed, and their resulting CCS values, could provide information as to the orientations and structures of the peptides while housed within the ND. Figure IV-4 shows the CCS values recorded for rIAPP dimer 4^+ bound to between 0 – 10 DMPC lipids. The CCS values obtained for IAPP-lipid complexes increase as a function of the number of lipids bound to rIAPP dimer. While the overall growth trend for the IAPP-lipid CCS data appears linear, we obtain significantly improved correlation coefficients when two linear regression models are used

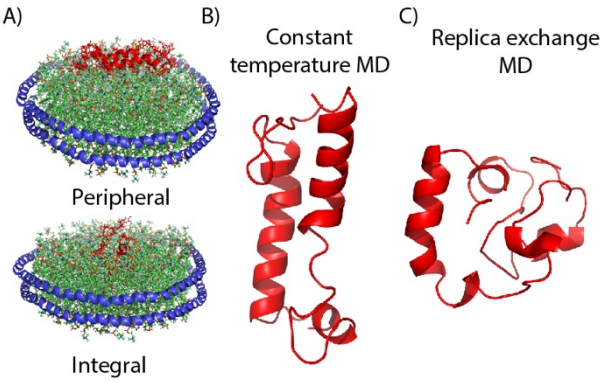


Appendix: Figure IV-4. CCS as a function of the number of DMPC lipids for rIAPP dimer 4^+ ions (black) and gramicidin A dimer 3^+ ions (red). Two linear regression models are required to fit the data for rIAPP dimer due to the systematic error seen in the residuals (inset).

to fit the data (Inset in Figure IV-4), indicating a shift in IAPP dimer structure upon binding either 6 or 7 DMPC lipids. Control experiments using Gramicidin A dimer 3^+ ions also exhibit a linear increase in CCS as a function of the number of lipids incorporated in the complex (Figure IV-4), but these data also produce a steeper slope, fewer lipid bound states, and optimal fit to a single linear function, in contrast to our IAPP data. The differences detected in Figure IV-4 between IAPP and gramicidin A dimers reflect differences in the binding mode accessed by these peptides when interacting with NDs. Gramicidin A dimer is known to form a transmembrane channel.[318] As such, differences between the CCS data recorded for gramicidin A and IAPP in the presences of NDs support the conclusion that the two peptides interact with cellular membranes differently, suggesting a peripheral interaction mode for the latter peptide.

To assign structural models for the data shown in Figure IV-4, we undertook a series of MD simulations in order to construct model structures and compare their CCS values to those observed experimentally. A previous report[324] has described IM-MS data and model structures for gramicidin A dimer associated with NDs. This report includes a CCS value for a single-stranded head-to-head (SSHH) gramicidin A dimer, relaxed in the gas-phase, to be 657±23 $Å^2$, a value that agrees within experimental uncertainty (5%) with our measured CCS values for gramicidin A dimer ions. This indicates that the structure of gramicidin A dimer liberated from ND likely retains the SSHH conformation adopted in its integrated, transmembrane configuration adopted within NDs. On the other hand, our results suggest that rIAPP dimers undergo significant structural collapse in the gas phase. The vacuum-relaxed rIAPP dimer structures from our constant temperature MD simulations produce CCS values of 1078±8 $Å^2$ and 1160±11 $Å^2$ for models peripherally and integrally associated with NDs, respectively (Figure IV-5). These values are significantly greater than those observed experimentally (11% and 19%, respectively). This result

led us to perform replica exchange MD simulations on rIAPP dimers in order to more thoroughly explore the most thermodynamically stable conformations of rIAPP dimer ions in the gas phase. These simulations produced globular type structures for rIAPP dimers exhibiting CCS values centering on 1137 ±18 Å^2 (Figure IV-5), a value that remains 17% greater than our experimental CCS values for rIAPP dimer ions. Taken together, our MD results strongly suggest that rIAPP dimer undergoes significant compaction in the gas phase in a manner inconsistent with integral binding within a ND. As such, these findings suggest that due its peripheral binding to NDs, rIAPP dimer is unshielded from the gas-phase and can undergo facile rearrangements on the ND surface. This situation is contrasted in our gramicidin A data, where the peptide remains protected and thus retains a stronger memory of its structure within the ND lipid bilayer. This result is supported by previous data[314] indicating that IAPP aggregates generally interact on the outer leaflet of the membrane bilayer, where some evidence of partial insertion into the membrane was also detected. Higher stoichiometry rIAPP dimer-lipid complexes exhibit a steeper slope in Figure IV-2, which can be interpreted as evidence of an alternative structural state consistent with partial insertion into the ND. Overall, our IM-MS analysis and computational modeling show that rIAPP dimer mainly interacts with the membrane in a peripheral mode.



Appendix: Figure IV-5. Models of rIAPP dimer. A) and B) show the peripheral and integral placement of rIAPP dimer in the ND. B) A representative rIAPP dimer structure from constant temperature MD simulation. C) A representative rIAPP dimer structure from replica exchange MD simulation.

Conclusions

The study of the mechanism of interaction between IAPP aggregates and membrane bilayer is important as it could provide insights into IAPP cytotoxicity and T2DM. We used IM-MS and computational modeling to study the interactions between IAPP and NDs. Our IM-MS observations show that higher-order oligomeric species of rIAPP were present in the ND. When ejected, rIAPP oligomeric species and lipid-bound states were identified, including peptide complexes bound to up to 10 lipids. When compared with control data acquired for gramicidin A, we observe different CCS growth trends and lipid stoichiometries rIAPP complexes released from NDs, strongly suggesting a peripheral binding mode for the complex. This conclusion is further supported from MD simulations, where greater degrees of structural rearrangement are observed for rIAPP than for gramicidin A. As the latter peptide dimer is known to be protected in an integral, transmembrane configuration within NDs, the structural rearrangements we detect are likely driven by peripheral binding to the surface of NDs. Future work will seek to extend our IM-MS methods to investigate a wider array of peptides, and drug interactions, in an effort to deduce greater functional details from peptide-ND interactions.

Bibliography

1. Osadchy, M. & Kolodny, R. Maps of protein structure space reveal a fundamental relationship between protein structure and function. *Proc. Natl. Acad. Sci.* **108**, 12301–12306 (2011).
2. Sutto, L., Marsili, S., Valencia, A. & Gervasio, F. L. From residue coevolution to protein conformational ensembles and functional dynamics. *Proc Natl Acad Sci U S A* **112**, 13567–13572 (2015).
3. Wang, R. Y.-R. *et al.* De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **12**, 335–338 (2015).
4. Bai, X., McMullan, G. & Scheres, S. H. . How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
5. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
6. Yang, H., Yang, S., Kong, J., Dong, A. & Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* **10**, 382–396 (2015).
7. Opella, S. J. & Marassi, F. M. Applications of NMR to membrane proteins. *Arch. Biochem. Biophys.* **628**, 92–101 (2017).
8. Andreas, L. B. *et al.* Structure of fully protonated proteins by proton-detected magic-Angle spinning NMR. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9187–9192 (2016).
9. Różycki, B. & Boura, E. Large, dynamic, multi-protein complexes: a challenge for structural biology. *J. Phys. Condens. Matter* **26**, 463103 (2014).
10. Feig, M., Yu, I., Wang, P., Nawrocki, G. & Sugita, Y. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *J. Phys. Chem. B* **121**, 8009–8025 (2017).
11. Popot, J.-L. Membrane Proteins and Their Natural Environment. in 1–57 (Springer, Cham, 2018). doi:10.1007/978-3-319-73148-3_1
12. KENDREW, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666 (1958).
13. Burley, S. K. *et al.* Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. in 627–641 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-7000-1_26
14. Morales-Rios, E., Montgomery, M. G., Leslie, A. G. W. & Walker, J. E. Structure of ATP synthase from Paracoccus denitrificans determined by X-ray crystallography at 4.0 Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13231–6 (2015).
15. Gres, A. T. *et al.* X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99–104 (2015).
16. Zheng, H. *et al.* X-ray crystallography over the past decade for novel drug discovery – where are we heading next? *Expert Opin. Drug Discov.* **10**, 975–989 (2015).
17. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of

membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18**, 581–586 (2008).
18. Schiebel, J. *et al.* High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **24**, 1398–1409 (2016).
19. Broecker, J. *et al.* A Versatile System for High-Throughput In Situ X-ray Screening and Data Collection of Soluble and Membrane-Protein Crystals. *Cryst. Growth Des.* **16**, 6318–6326 (2016).
20. Schieferstein, J. M. *et al.* X-ray transparent microfluidic chips for high-throughput screening and optimization of *in meso* membrane protein crystallization. *Biomicrofluidics* **11**, 024118 (2017).
21. Rucktooa, P. *et al.* Towards high throughput GPCR crystallography: In Meso soaking of Adenosine A2A Receptor crystals. *Sci. Rep.* **8**, 41 (2018).
22. Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci.* **104**, 9615–9620 (2007).
23. Tang, Y. *et al.* Protein structure determination by combining sparse NMR data with evolutionary couplings. *Nat. Methods* **12**, 751–754 (2015).
24. Pan, B.-B. *et al.* 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy. *Chem. Commun.* **52**, 10237–10240 (2016).
25. Quinn, C. M. & Polenova, T. Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Q. Rev. Biophys.* **50**, e1 (2017).
26. Bonomi, M., Heller, G. T., Camilloni, C. & Vendruscolo, M. Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **42**, 106–116 (2017).
27. Kovermann, M., Rogne, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6 (2016).
28. Lange, O. F. *et al.* Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10873–8 (2012).
29. Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–8 (2010).
30. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **328**, 98–102 (2010).
31. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **17**, 768–774 (2010).
32. Shahid, S. A. *et al.* Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nat. Methods* **9**, 1212–1217 (2012).
33. Wang, S. *et al.* Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat. Methods* **10**, 1007–1012 (2013).
34. Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. & Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* **278**, 687–703 (2011).
35. Fernandez-Leiro, R. & Scheres, S. H. W. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **537**, 339–346 (2016).
36. DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).
37. Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam, S. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11709–14 (2014).

38. Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *Elife* **4**, (2015).
39. Franken, L. E. *et al.* A general mechanism of ribosome dimerization revealed by single-particle cryo-electron microscopy. *Nat. Commun.* **8**, 722 (2017).
40. Bai, X., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2**, (2013).
41. Zubcevic, L. *et al.* Cryo-electron microscopy structure of the TRPV2 ion channel. *Nat. Struct. Mol. Biol.* **23**, 180–186 (2016).
42. Jin, P. *et al.* Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature* **547**, 118–122 (2017).
43. Winkler, P. A., Huang, Y., Sun, W., Du, J. & Lü, W. Electron cryo-microscopy structure of a human TRPM4 channel. *Nature* **552**, 200–204 (2017).
44. Yang, S. Methods for SAXS-Based Structure Determination of Biomolecular Complexes. *Adv. Mater.* **26**, 7902–7910 (2014).
45. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141 (2010).
46. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3095-103 (2015).
47. Miles, A. J. & Wallace, B. A. Circular dichroism spectroscopy of membrane proteins. *Chem. Soc. Rev.* **45**, 4859–4872 (2016).
48. Titus, J., Ghimire, H., Viennois, E., Merlin, D. & Unil Perera, A. G. Protein secondary structure analysis of dried blood serum using infrared spectroscopy to identify markers for colitis screening. *J. Biophotonics* **11**, e201700057 (2018).
49. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* 1–17 (2019). doi:10.1038/s41580-019-0163-x
50. Lee, J., Freddolino, P. L. & Zhang, Y. Ab Initio Protein Structure Prediction. in *From Protein Structure to Function with Bioinformatics* 3–35 (Springer Netherlands, 2017). doi:10.1007/978-94-024-1069-3_1
51. Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)-Round XII. *Proteins Struct. Funct. Bioinforma.* **86**, 7–15 (2018).
52. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **7**, 10480 (2017).
53. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).
54. Lopes, P. E. M., Guvench, O. & MacKerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics Simulations. in 47–71 (Humana Press, New York, NY, 2015). doi:10.1007/978-1-4939-1465-4_3
55. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4758–E4766 (2018).
56. Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
57. Margreitter, C., Reif, M. M. & Oostenbrink, C. Update on phosphate and charged post-

translationally modified amino acid parameters in the GROMOS force field. *J. Comput. Chem.* **38**, 714–720 (2017).
58. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
59. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
60. Vanommeslaeghe, K. & MacKerell, A. D. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 861–871 (2015).
61. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
62. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
63. Yang, J. & Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. in *Current Protocols in Bioinformatics* **52**, 5.8.1-5.8.15 (John Wiley & Sons, Inc., 2015).
64. Xu, D. & Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins Struct. Funct. Bioinforma.* **80**, n/a-n/a (2012).
65. Xu, D. & Zhang, Y. Toward optimal fragment generations for *ab initio* protein structure assembly. *Proteins Struct. Funct. Bioinforma.* **81**, 229–239 (2013).
66. Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 872–877 (2015).
67. Webb, B. *et al.* Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018).
68. Heuvel, R. H. van den & Heck, A. J. Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**, 519–526 (2004).
69. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008).
70. Benesch, J. L. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21**, 641–649 (2011).
71. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 1239–1256 (2013).
72. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* **23**, 252–260 (2013).
73. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **9**, 449–472 (2016).
74. Cannon, J. *et al.* High-Throughput Middle-Down Analysis Using an Orbitrap. *J. Proteome Res.* **9**, 3886–3890 (2010).
75. Sidoli, S. & Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev. Proteomics* **14**, 617–626 (2017).
76. Fornelli, L., Ayoub, D., Aizikov, K., Beck, A. & Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **86**, 3005–3012 (2014).

77. Sweredoski, M. J., Moradian, A., Raedle, M., Franco, C. & Hess, S. High Resolution Parallel Reaction Monitoring with Electron Transfer Dissociation for Middle-Down Proteomics. *Anal. Chem.* **87**, 8360–8366 (2015).
78. Toby, T. K., Fornelli, L. & Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519 (2016).
79. Cai, W. *et al.* Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal. Chem.* **89**, 5467–5475 (2017).
80. Fornelli, L. *et al.* Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **16**, 609–618 (2017).
81. Li, K. S., Shi, L. & Gross, M. L. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc. Chem. Res.* **51**, 736–744 (2018).
82. Wang, L. & Chance, M. R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol. Cell. Proteomics* **16**, 706–716 (2017).
83. Heinkel, F. & Gsponer, J. Determination of Protein Folding Intermediate Structures Consistent with Data from Oxidative Footprinting Mass Spectrometry. *J. Mol. Biol.* **428**, 365–371 (2016).
84. Michael J. Chalmers, † *et al.* Probing Protein Ligand Interactions by Automated Hydrogen/Deuterium Exchange Mass Spectrometry. (2006). doi:10.1021/AC051294F
85. Houde, D., Arndt, J., Domeier, W., Berkowitz, S. & Engen, J. R. Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **81**, 2644–2651 (2009).
86. Wei, H. *et al.* Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug Discov. Today* **19**, 95–102 (2014).
87. Kochert, B. A., Iacob, R. E., Wales, T. E., Makriyannis, A. & Engen, J. R. Hydrogen-Deuterium Exchange Mass Spectrometry to Study Protein Complexes. in 153–171 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-7759-8_10
88. Pan, J., Han, J., Borchers, C. H. & Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 kDa Protein. *J. Am. Chem. Soc.* **131**, 12801–12808 (2009).
89. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **25**, 663–682 (2006).
90. Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9**, 1634–49 (2010).
91. Arlt, C., Ihling, C. H. & Sinz, A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. *Proteomics* n/a-n/a (2015). doi:10.1002/pmic.201400549
92. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
93. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* **41**, 20–32 (2016).
94. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008).
95. Boeri Erba, E. & Petosa, C. The emerging role of native mass spectrometry in

characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* **24**, 1176–1192 (2015).
96. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
97. Li, H., Nguyen, H. H., Ogorzalek Loo, R. R., Campuzano, I. D. G. & Loo, J. A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139–148 (2018).
98. Ro, S. Y. *et al.* Native top-down mass spectrometry provides insights into the copper centers of membrane-bound methane monooxygenase. *Nat. Commun.* **10**, 2675 (2019).
99. Smith, R. D., Loo, J. A., Loo, R. R. O., Busman, M. & Udseth, H. R. Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* **10**, 359–452 (1991).
100. Light-Wahl, K. J., Schwartz, B. L. & Smith, R. D. Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **116**, 5271–5278 (1994).
101. Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1–23 (1997).
102. Li, J. *et al.* Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrom. Rev.* **35**, 111–122 (2016).
103. McAllister, R. G., Metwally, H., Sun, Y. & Konermann, L. Release of Native-like Gaseous Proteins from Electrospray Droplets via the Charged Residue Mechanism: Insights from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **137**, 12667–12676 (2015).
104. Kim, D., Wagner, N., Wooding, K., Clemmer, D. E. & Russell, D. H. Ions from Solution to the Gas Phase: A Molecular Dynamics Simulation of the Structural Evolution of Substance P during Desolvation of Charged Nanodroplets Generated by Electrospray Ionization. *J. Am. Chem. Soc.* **139**, 2981–2988 (2017).
105. Mitchell Wells, J. & McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **402**, 148–185 (2005).
106. Armirotti, A., Benatti, U. & Damonte, G. Top-down proteomics with a quadrupole time-of-flight mass spectrometer and collision-induced dissociation. *Rapid Commun. Mass Spectrom.* **23**, 661–666 (2009).
107. Liu, F. & Goshe, M. B. Combinatorial Electrostatic Collision-Induced Dissociative Chemical Cross-linking Reagents for Probing Protein Surface Topology. *Anal. Chem.* **82**, 6215–6223 (2010).
108. Popa, V., Trecroce, D. A., McAllister, R. G. & Konermann, L. Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. *J. Phys. Chem. B* **120**, 5114–5124 (2016).
109. Song, Y., Nelp, M. T., Bandarian, V. & Wysocki, V. H. Refining the Structural Model of a Heterohexameric Protein Complex: Surface Induced Dissociation and Ion Mobility Provide Key Connectivity and Topology Information. *ACS Cent. Sci.* **1**, 477–487 (2015).
110. Busch, F. *et al.* Localization of Protein Complex Bound Ligands by Surface-Induced Dissociation High-Resolution Mass Spectrometry. *Anal. Chem.* **90**, 12796–12801 (2018).
111. Harvey, S. R. *et al.* Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8143–8148 (2019).

112. Sahasrabuddhe, A. *et al.* Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1268–1273 (2018).
113. Lermyte, F. & Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **15**, 2813–2822 (2015).
114. Riley, N. M., Westphall, M. S. & Coon, J. J. Sequencing Larger Intact Proteins (30-70 kDa) with Activated Ion Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **29**, 140–149 (2018).
115. Schennach, M. & Breuker, K. Probing Protein Structure and Folding in the Gas Phase by Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **26**, 1059–1067 (2015).
116. Wongkongkathep, P. *et al.* Enhancing protein disulfide bond cleavage by UV excitation and electron capture dissociation for top-down mass spectrometry. *Int. J. Mass Spectrom.* **390**, 137–145 (2015).
117. Zhang, Y. *et al.* Native MS and ECD Characterization of a Fab–Antigen Complex May Facilitate Crystallization for X-ray Diffraction. *J. Am. Soc. Mass Spectrom.* **27**, 1139–1142 (2016).
118. Zhang, H., Cui, W., Wen, J., Blankenship, R. E. & Gross, M. L. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* **83**, 5598–606 (2011).
119. Warnke, S., von Helden, G. & Pagel, K. Analyzing the higher order structure of proteins with conformer-selective ultraviolet photodissociation. *Proteomics* **15**, 2804–2812 (2015).
120. Cammarata, M. B., Thyer, R., Rosenberg, J., Ellington, A. & Brodbelt, J. S. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **137**, 9128–9135 (2015).
121. Morrison, L. J. & Brodbelt, J. S. 193 nm Ultraviolet Photodissociation Mass Spectrometry of Tetrameric Protein Complexes Provides Insight into Quaternary and Secondary Protein Topology. *J. Am. Chem. Soc.* **138**, 10849–10859 (2016).
122. Hendricks, N. G. & Julian, R. R. Leveraging ultraviolet photodissociation and spectroscopy to investigate peptide and protein three-dimensional structure with mass spectrometry. *Analyst* **141**, 4534–4540 (2016).
123. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* **6**, 281–294 (2014).
124. Allen, S. J., Giles, K., Gilbert, T. & Bush, M. F. Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst* **141**, 884–891 (2016).
125. May, J. C. & McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **87**, 1422–1436 (2015).
126. Ben-Nissan, G. & Sharon, M. The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes. *Curr. Opin. Chem. Biol.* **42**, 25–33 (2018).
127. Gabelica, V. & Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Opin. Chem. Biol.* **42**, 51–59 (2018).
128. Mason, E. A. & Schamp, H. W. Mobility of gaseous lons in weak electric fields. *Ann. Phys. (N. Y).* **4**, 233–270 (1958).

129. Bleiholder, C., Dupuis, N. F., Wyttenbach, T. & Bowers, M. T. Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. *Nat. Chem.* **3**, 172–177 (2011).
130. Smith, D. P., Radford, S. E. & Ashcroft, A. E. Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6794–8 (2010).
131. Bernstein, S. L. *et al.* Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326–331 (2009).
132. Bernstein, S. L. *et al.* α-Synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation. *J. Am. Soc. Mass Spectrom.* **15**, 1435–1443 (2004).
133. Vlad, C. *et al.* Autoproteolytic Fragments Are Intermediates in the Oligomerization/Aggregation of the Parkinson's Disease Protein Alpha-Synuclein as Revealed by Ion Mobility Mass Spectrometry. *ChemBioChem* **12**, 2740–2744 (2011).
134. Woods, L. A., Radford, S. E. & Ashcroft, A. E. Advances in ion mobility spectrometry–mass spectrometry reveal key insights into amyloid assembly. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 1257–1268 (2013).
135. Soper, M. T., DeToma, A. S., Hyung, S.-J., Lim, M. H. & Ruotolo, B. T. Amyloid-β–neuropeptide interactions assessed by ion mobility-mass spectrometry. *Phys. Chem. Chem. Phys.* **15**, 8952 (2013).
136. Ghassabi Kondalaji, S., Khakinejad, M., Tafreshian, A. & J. Valentine, S. Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 1. An Advanced Protocol for Molecular Dynamics Simulations and Collision Cross-Section Calculation. *J. Am. Soc. Mass Spectrom.* **28**, 947–959 (2017).
137. Marklund, E. G., Degiacomi, M. T., Robinson, C. V, Baldwin, A. J. & Benesch, J. L. Collision cross sections for structural proteomics. *Structure* **23**, 791–799 (2015).
138. Larriba, C. & Hogan, C. J. Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **117**, 3887–3901 (2013).
139. Larriba, C. & Hogan, C. J. Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation. *J. Comput. Phys.* **251**, 344–363 (2013).
140. Politis, A. *et al.* Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS One* **5**, e12080 (2010).
141. Hall, Z., Politis, A. & Robinson, C. V. Structural Modeling of Heteromeric Protein Complexes from Disassembly Pathways and Ion Mobility-Mass Spectrometry. *Structure* **20**, 1596–1609 (2012).
142. Hall, Z., Politis, A. & Robinson, C. V. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **20**, 1596–1609 (2012).
143. Politis, A., Park, A. Y., Hall, Z., Ruotolo, B. T. & Robinson, C. V. Integrative modelling coupled with ion mobility mass spectrometry reveals structural features of the clamp loader in complex with single-stranded DNA binding protein. *J Mol Biol* **425**, 4790–4801 (2013).
144. Politis, A. *et al.* Topological Models of Heteromeric Protein Assemblies from Mass Spectrometry: Application to the Yeast eIF3:eIF5 Complex. *Chem. Biol.* **22**, 117–128

(2015).
145. Eschweiler, J. D., Farrugia, M. A., Dixit, S. M., Hausinger, R. P. & Ruotolo, B. T. A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure* **26**, 599-606.e3 (2018).
146. Eschweiler, J. D., Frank, A. T. & Ruotolo, B. T. Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J. Am. Soc. Mass Spectrom.* **28**, 1991–2000 (2017).
147. Giles, K. *et al.* Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **18**, 2401–2414 (2004).
148. Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**, 1–12 (2007).
149. May, J. C., Morris, C. B. & Mclean, J. A. Ion Mobility Collision Cross Section Compendium. (2016). doi:10.1021/acs.analchem.6b04905
150. Deng, L. *et al.* Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS using Structures for Lossless Ion Manipulations. *Anal. Chem.* **89**, 4628–4634 (2017).
151. Giles, K., Williams, J. P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* **25**, 1559–1566 (2011).
152. Shvartsburg, A. A. & Smith, R. D. Fundamentals of traveling wave ion mobility spectrometry. *Anal. Chem.* **80**, 9689–99 (2008).
153. Richardson, K., Langridge, D. & Giles, K. Fundamentals of travelling wave ion mobility revisited: I. Smoothly moving waves. *Int. J. Mass Spectrom.* **428**, 71–80 (2018).
154. Wyttenbach, T., Von Helden, G. & Bowers, M. T. Gas-phase conformation of biological molecules: Bradykinin. *J. Am. Chem. Soc.* **118**, 8355–8364 (1996).
155. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. Ion mobility–mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**, 1139–1152 (2008).
156. Mortensen, D. N., Susa, A. C. & Williams, E. R. Collisional Cross-Sections with T-Wave Ion Mobility Spectrometry without Experimental Calibration. *J. Am. Soc. Mass Spectrom.* **28**, 1282–1292 (2017).
157. Dixit, S. M., Polasky, D. A. & Ruotolo, B. T. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr. Opin. Chem. Biol.* **42**, 93–100 (2018).
158. Shelimov, K. B. & Jarrold, M. F. Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **119**, 2987–2994 (1997).
159. Niu, S. & Ruotolo, B. T. Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding. *Protein Sci.* **24**, 1272–1281 (2015).
160. Rabuck, J. N. *et al.* Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry. *Anal. Chem.* **85**, 6995–7002 (2013).
161. Zhong, Y., Han, L. & Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie Int. Ed.* **53**, 9209–9212 (2014).
162. Beveridge, R. *et al.* Mass spectrometry locates local and allosteric conformational changes that occur on cofactor binding. *Nat. Commun.* **7**, 12163 (2016).

163. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–175 (2014).
164. Tian, Y., Han, L., Buckner, A. C. & Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **87**, 11509–11515 (2015).
165. Hernandez-Alba, O., Wagner-Rousset, E., Beck, A. & Cianférani, S. Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding for Conformational Characterization of IgG4 Monoclonal Antibodies. *Anal. Chem.* **90**, 8865–8872 (2018).
166. Dong, S., Wagner, N. D. & Russell, D. H. Collision-Induced Unfolding of Partially Metalated Metallothionein-2A: Tracking Unfolding Reactions of Gas-Phase Ions. *Anal. Chem.* **90**, 11856–11862 (2018).
167. Rabuck-Gibbons, J. N., Lodge, J. M., Mapp, A. K. & Ruotolo, B. T. Collision-Induced Unfolding Reveals Unique Fingerprints for Remote Protein Interaction Sites in the KIX Regulation Domain. *J. Am. Soc. Mass Spectrom.* **30**, 94–102 (2019).
168. Tian, Y. & Ruotolo, B. T. Collision induced unfolding detects subtle differences in intact antibody glycoforms and associated fragments. *Int. J. Mass Spectrom.* **425**, 1–9 (2018).
169. Rabuck-Gibbons, J. N., Keating, J. E. & Ruotolo, B. T. Collision induced unfolding and dissociation differentiates ATP-competitive from allosteric protein tyrosine kinase inhibitors. *Int. J. Mass Spectrom.* **427**, 151–156 (2018).
170. Tian, Y. & Ruotolo, B. T. Collision induced unfolding detects subtle differences in intact antibody glycoforms and associated fragments. *Int. J. Mass Spectrom.* **425**, 1–9 (2018).
171. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
172. van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 6589–6631 (2014).
173. Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553–584 (2014).
174. Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science (80-.).* **320**, 1471–1475 (2008).
175. Bernado, P., Mylonas, E., Petoukhov, M. V, Blackledge, M. & Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* **129**, 5656–5664 (2007).
176. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **127**, 476–477 (2005).
177. Nakajima, N., Nakamura, H. & Kidera, A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **101**, 817–824 (1997).
178. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chemie-International Ed.* **45**, 4064–4092 (2006).
179. Kelleher, N. L. From primary structure to function: Biological insights from large-molecule mass spectra. *Chem. Biol.* **7**, 37–45 (2000).
180. van den Heuvel, R. H. & Heck, A. J. R. Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**, 519–526 (2004).
181. Mehmood, S. *et al.* Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24. *Nat. Chem.* **8**, 1152–

1158 (2016).
182. Chan, D. S.-H. *et al.* Effect of DMSO on Protein Structure and Interactions Assessed by Collision-Induced Dissociation and Unfolding. *Anal. Chem.* **89**, 9976–9983 (2017).
183. Zhang, Z., Browne, S. J. & Vachet, R. W. Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J Am Soc Mass Spectrom* **25**, 604–613 (2014).
184. O'Brien, J. P., Li, W. Z., Zhang, Y. & Brodbelt, J. S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **136**, 12920–12928 (2014).
185. Mistarz, U. H., Brown, J. M., Haselmann, K. F. & Rand, K. D. Probing the Binding Interfaces of Protein Complexes Using Gas-Phase H/D Exchange Mass Spectrometry. *Structure* **24**, 310–318 (2016).
186. Zhang, A., Fang, J., Chou, R. Y., Bondarenko, P. V & Zhang, Z. Conformational difference in human IgG2 disulfide isoforms revealed by hydrogen/deuterium exchange mass spectrometry. *Biochemistry* **54**, 1956–1962 (2015).
187. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics* **13**, 2927–2943 (2014).
188. Konermann, L., Vahidi, S. & Sowole, M. A. Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem* **86**, 213–232 (2014).
189. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294 (2014).
190. May, J. C., Morris, C. B. & McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Anal Chem* **89**, 1032–1044 (2017).
191. Allen, S. J., Giles, K., Gilbert, T. & Bush, M. F. Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst* **141**, 884–891 (2016).
192. Politis, A., Park, A. Y., Hall, Z., Ruotolo, B. T. & Robinson, C. V. Integrative modelling coupled with ion mobility mass spectrometry reveals structural features of the clamp loader in complex with single-stranded DNA binding protein. *J. Mol. Biol.* **425**, 4790–4801 (2013).
193. Eschweiler, J. D., Frank, A. T. & Ruotolo, B. T. Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J Am Soc Mass Spectrom* (2017). doi:10.1007/s13361-017-1757-1
194. Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. & Jarrold, M. F. Protein structure in Vacuo: Gas-phase conformations of BPTI and cytochrome c. *J. Am. Chem. Soc.* **119**, 2240–2248 (1997).
195. Shi, H., Pierson, N. a., Valentine, S. J. & Clemmer, D. E. Conformation types of ubiquitin [M+8H] 8+ ions from water:methanol solutions: Evidence for the N and A states in aqueous solution. *J. Phys. Chem. B* **116**, 3344–3352 (2012).
196. Hudgins, R. R., Mao, Y., Ratner, M. A. & Jarrold, M. F. Conformations of Gly(n)H+ and Ala(n)H+ peptides in the gas phase. *Biophys. J.* **76**, 1591–7 (1999).
197. Giles, K. *et al.* Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401–2414 (2004).
198. Kune, C., Far, J. & De Pauw, E. Accurate Drift Time Determination by Traveling Wave

Ion Mobility Spectrometry: The Concept of the Diffusion Calibration. *Anal Chem* **88**, 11639–11646 (2016).
199. Mortensen, D. N., Susa, A. C. & Williams, E. R. Collisional Cross-Sections with T-Wave Ion Mobility Spectrometry without Experimental Calibration. *J Am Soc Mass Spectrom* **28**, 1282–1292 (2017).
200. Manura, D. & Dahl, D. *SIMION (R) 8.0/8.1 User Manual*. (Scientific Instrument Services, 2011).
201. Buck, M., Bouguet-Bonnet, S., Pastor, R. W. & MacKerell Jr., A. D. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys J* **90**, L36-8 (2006).
202. Mullner, D. Modern hierarchical, agglomerative clustering algorithms. *eprint arXiv:1109.2378* (2011).
203. Zepeda-Mendoza, M. L. & Resendis-Antonio, O. Hierarchical Agglomerative Clustering. in *Encyclopedia of Systems Biology* (eds. Dubitzky, W., Wolkenhauer, O., Cho, K.-H. & Yokota, H.) 886–887 (Springer New York, 2013). doi:10.1007/978-1-4419-9863-7_1371
204. Walt, S. van der, Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
205. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20 (2007).
206. Haynes, S. E. *et al.* Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Anal. Chem.* **89**, 5669–5672 (2017).
207. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 99–104 (2007).
208. Mason, E. A. & McDaniel, E. W. *Transport properties of ions in gases*. (Wiley, 1988).
209. Siems, W. F. *et al.* Measuring the Resolving Power of Ion Mobility Spectrometers. *Anal. Chem.* **66**, 4195–4201 (1994).
210. Wu, C., Siems, W. F., Asbury, G. R. & Hill, H. H. Electrospray ionization high-resolution ion mobility spectrometry - Mass spectrometry. *Anal. Chem.* **70**, 4929–4938 (1998).
211. Zhong, Y., Hyung, S.-J. & Ruotolo, B. T. Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *Analyst* **136**, 3534 (2011).
212. Hudgins, R. R., Ratner, M. A. & Jarrold, M. F. Design of helices that are stable in vacuo. *J. Am. Chem. Soc.* **120**, 12974–12975 (1998).
213. Hudgins, R. R. & Jarrold, M. F. Helix formation in unsolvated alanine-based peptides: Helical monomers and helical dimers. *J. Am. Chem. Soc.* **121**, 3494–3501 (1999).
214. Giles, K., Williams, J. P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* **25**, 1559–1566 (2011).
215. Dodds, J. N., May, J. C. & McLean, J. A. Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Anal. Chem.* **89**, 12176–12184 (2017).
216. Morsa, D., Gabelica, V. & De Pauw, E. Effective temperature of ions in traveling wave ion mobility spectrometry. *Anal Chem* **83**, 5775–5782 (2011).
217. Merenbloom, S. I., Flick, T. G. & Williams, E. R. How hot are your ions in TWAVE ion mobility spectrometry? *J Am Soc Mass Spectrom* **23**, 553–562 (2012).
218. Swaim, C. L., Smith, J. B. & Smith, D. L. Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides.

J. Am. Soc. Mass Spectrom. **15**, 736–749 (2004).
219. Clifford-Nunn, B., Showalter, H. D. H. & Andrews, P. C. Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions. *J. Am. Soc. Mass Spectrom.* **23**, 201–212 (2012).
220. Rozbeský, D. *et al.* Impact of Chemical Cross-Linking on Protein Structure and Function. *Anal. Chem.* **90**, 1104–1113 (2018).
221. St Louis, R. H., Hill Jr, H. H., Alan Eiceman, G., Hill, H. H. & Referee, J. Ion Mobility Spectrometry in Analytical Chemistry. *Crit. Rev. Anal. Chem.* **21**, 321–355 (1990).
222. Ewing, R. G., Atkinson, D. A., Eiceman, G. A. & Ewing, G. J. A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds. *Talanta* **54**, 515–529 (2001).
223. Puton, J. & Namieśnik, J. Ion mobility spectrometry: Current status and application for chemical warfare agents detection. *TrAC Trends Anal. Chem.* **85**, 10–20 (2016).
224. Hernández-Mesa, M., Escourrou, A., Monteau, F., Le Bizec, B. & Dervilly-Pinel, G. Current applications and perspectives of ion mobility spectrometry to answer chemical food safety issues. *TrAC Trends Anal. Chem.* **94**, 39–53 (2017).
225. May, J. C. *et al.* Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **86**, 2107–2116 (2014).
226. Dodds, J. N. & Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* 1–11 (2019). doi:10.1007/s13361-019-02288-2
227. Beata M. Kolakowski, †, Paul A. D'Agostino, ‡, Claude Chenier, ‡ and & Zoltán Mester*, †. Analysis of Chemical Warfare Agents in Food Products by Atmospheric Pressure Ionization-High Field Asymmetric Waveform Ion Mobility Spectrometry-Mass Spectrometry. (2007). doi:10.1021/AC070816J
228. McLean, J. A., Ridenour, W. B. & Caprioli, R. M. Profiling and imaging of tissues by imaging ion mobility-mass spectrometry. *J. Mass Spectrom.* **42**, 1099–1105 (2007).
229. McLean, J. A., Ruotolo, B. T., Gillig, K. J. & Russell, D. H. Ion mobility–mass spectrometry: a new paradigm for proteomics. *Int. J. Mass Spectrom.* **240**, 301–315 (2005).
230. Helm, D. *et al.* Ion Mobility Tandem Mass Spectrometry Enhances Performance of Bottom-up Proteomics □ S. *Mol. Cell. Proteomics* **13**, 3709–3715 (2014).
231. Zhang, X., Quinn, K., Cruickshank-Quinn, C., Reisdorph, R. & Reisdorph, N. The application of ion mobility mass spectrometry to metabolomics. *Curr. Opin. Chem. Biol.* **42**, 60–66 (2018).
232. Paglia, G. & Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **12**, (2017).
233. El-baba, T. J. *et al.* Melting Proteins: Evidence for Multiple Stable Structures upon Thermal Denaturation of Native Ubiquitin from IMS-MS Measurements. *J. Am. Chem. Soc.* jacs.7b02774 (2017). doi:10.1021/jacs.7b02774
234. Politis, A., Park, A. Y., Hyung, S.-J., Barsky, D. & Ruotolo, B. T. Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS One* **5**, 12080 (2010).
235. Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int.*

J. Mass Spectrom. **261**, 1–12 (2007).
236. Shvartsburg, A. A. & Smith, R. D. Fundamentals of traveling wave ion mobility spectrometry. *Anal Chem* **80**, 9689–9699 (2008).
237. Richardson, K., Langridge, D. & Giles, K. Fundamentals of travelling wave ion mobility revisited: I. Smoothly moving waves. *Int. J. Mass Spectrom.* **428**, 71–80 (2018).
238. Campuzano, I. *et al.* Structural Characterization of Drug-like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections. *Anal. Chem.* **84**, 1026–1033 (2012).
239. Hofmann, J. *et al.* Estimating Collision Cross Sections of Negatively Charged *N*- Glycans using Traveling Wave Ion Mobility-Mass Spectrometry. *Anal. Chem.* **86**, 10789–10795 (2014).
240. Hines, K. M., May, J. C., Mclean, J. A. & Xu, L. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. (2016). doi:10.1021/acs.analchem.6b01728
241. Lippens, J. L., Ranganathan, S. V, D'esposito, R. J. & Fabris, D. Modular calibrant sets for the structural analysis of nucleic acids by ion mobility spectrometry mass spectrometry †. *Analyst* **141**, 4084 (2016).
242. Bush, M. F., Campuzano, I. D. G. & Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem.* **84**, 7124–7130 (2012).
243. Bush, M. F. *et al.* Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **82**, 9557–9565 (2010).
244. Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* **3**, 1139–1152 (2008).
245. Dixit, S. M. & Ruotolo, B. T. A Semi-Empirical Framework for Interpreting Traveling Wave Ion Mobility Arrival Time Distributions. *J. Am. Soc. Mass Spectrom.* **30**, 956–966 (2019).
246. Haynes, S. E. *et al.* Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Anal. Chem.* **89**, (2017).
247. Giles, K. *et al.* A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **91**, 8564–8573 (2019).
248. Baker, E. S. *et al.* Enhancing bottom-up and top-down proteomic measurements with ion mobility separations. *Proteomics* 2766–2776 (2015). doi:10.1002/pmic.201500048
249. Meier, F. *et al.* Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).
250. Hinz, C., Liggi, S. & Griffin, J. L. The potential of Ion Mobility Mass Spectrometry for high-throughput and high-resolution lipidomics. *Curr. Opin. Chem. Biol.* **42**, 42–50 (2018).
251. Mairinger, T., Causon, T. J. & Hann, S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. *Curr. Opin. Chem. Biol.* **42**, 9–15 (2018).
252. Struwe, W. B., Pagel, K., Benesch, J. L. P., Harvey, D. J. & Campbell, M. P. GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics. doi:10.1007/s10719-015-9613-7
253. Picache, J. A. *et al.* Collision cross section compendium to annotate and predict multi-

omic compound identities. *Chem. Sci.* **10**, 983–993 (2019).
254. Hines, K. M., Ross, D. H., Davidson, K. L., Bush, M. F. & Xu, L. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Anal. Chem.* **89**, 9023–9030 (2017).
255. Paglia, G. *et al.* Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. *Anal. Chem.* **87**, 1137–1144 (2015).
256. Eschweiler, J. D., Farrugia, M. A., Dixit, S. M., Hausinger, R. P. & Ruotolo, B. T. A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure* **26**, 599-606.e3 (2018).
257. Gabelica, V. *et al.* Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrom. Rev.* **38**, 291–320 (2019).
258. Poincaré, H. & Magini, R. Les méthodes nouvelles de la mécanique céleste. *Nuovo Cim.* **10**, 128–130 (1899).
259. Silveira, J. A., Gamage, C. M., Blase, R. C. & Russell, D. H. Gas-phase ion dynamics in a periodic-focusing DC ion guide. *Int. J. Mass Spectrom.* **296**, 36–42 (2009).
260. Zhong, Y. Y., Hyung, S. J. & Ruotolo, B. T. Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *Analyst* **136**, 3534–3541 (2011).
261. Stow, S. M. *et al.* An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem.* **89**, 9048–9055 (2017).
262. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
263. Lössl, P., van de Waterbeemd, M. & Heck, A. J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J.* **35**, 2634–2657 (2016).
264. Mehmood, S., Allison, T. M. & Robinson, C. V. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* 453–474 (2015). doi:10.1146/annurev-physchem-040214-121732
265. Zhong, Y., Hyung, S.-J. & Ruotolo, B. T. Ion mobility–mass spectrometry for structural proteomics. *Expert Rev. Proteomics* **9**, 47–58 (2012).
266. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. (2014). doi:10.1038/NCHEM.1889
267. Terral, G., Beck, A. & Cianférani, S. Insights from native mass spectrometry and ion mobility-mass spectrometry for antibody and antibody-based product characterization. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1032**, 79–90 (2016).
268. Bleiholder, C. & Bowers, M. T. The Solution Assembly of Biological Molecules Using Ion Mobility Methods: From Amino Acids to Amyloid β-Protein. *Annu. Rev. Anal. Chem* **10**, 365–86 (2017).
269. Clemmer, D. E., Hudgins, R. R. & Jarrold, M. F. Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* **117**, 10141–10142 (1995).
270. Hopper, J. T. S. & Oldham, N. J. Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom.* **20**, 1851–1858 (2009).
271. Ruotolo, B. Searching for conformationally-selective small molecule therapeutics using ion mobility-mass spectrometry (227.1). *FASEB J.* **28**, (2014).
272. Pisupati, K. *et al.* A Multidimensional Analytical Comparison of Remicade and the

Biosimilar Remsima. *Anal. Chem.* acs.analchem.6b04436 (2017). doi:10.1021/acs.analchem.6b04436
273. Hernandez-Alba, O., Wagner-Rousset, E., Beck, A. & Cianférani, S. Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding for Conformational Characterization of IgG4 Monoclonal Antibodies. *Anal. Chem.* **90**, 8865–8872 (2018).
274. Watanabe, Y. *et al.* Signature of Antibody Domain Exchange by Native Mass Spectrometry and Collision-Induced Unfolding. *Anal. Chem.* **90**, 7325–7331 (2018).
275. Tian, Y., Lippens, J. L., Netirojjanakul, C., Campuzano, I. D. G. & Ruotolo, B. T. Quantitative collision-induced unfolding differentiates model antibody–drug conjugates. *Protein Sci.* **28**, 598–608 (2019).
276. Beveridge, R. *et al.* Mass spectrometry locates local and allosteric conformational changes that occur on cofactor binding. *Nat. Commun.* **7**, 12163 (2016).
277. Vallejo, D. D. *et al.* A Modified Drift Tube Ion Mobility-Mass Spectrometer for Charge Multiplexed Collision Induced Unfolding. *Under Rev.*
278. Li, F., Xia, Y., Meiler, J. & Ferguson-Miller, S. Characterization and modeling of the oligomeric state and ligand binding behavior of purified translocator protein 18 kDa from Rhodobacter sphaeroides. *Biochemistry* **52**, 5884–5899 (2013).
279. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–651 (2013).
280. Seeliger, M. A. *et al.* High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci.* **14**, 3135–3139 (2005).
281. Dowdy, S. M., Wearden, S. & Chilko, D. M. *Statistics for research*. *Wiley series in probability and statistics* (Wiley-Interscience, 2004).
282. Arlot, S. & Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010).
283. Izenman, A. J. Linear Discriminant Analysis. in 237–280 (Springer, New York, NY, 2013). doi:10.1007/978-0-387-78189-1_8
284. Izenman, A. J. Support Vector Machines. in 369–406 (Springer, New York, NY, 2013). doi:10.1007/978-0-387-78189-1_11
285. Liu, Y., Cong, X., Liu, W. & Laganowsky, A. Characterization of Membrane Protein–Lipid Interactions by Mass Spectrometry Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 579–586 (2017).
286. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nat. Commun.* **6**, 8551 (2015).
287. Fantin, S. M. *et al.* CIU classifies ligand binding behavior of integral membrane translocator protein TSPO. *Press*
288. Brown, M. T. & Cooper, J. A. Regulation, substrates and functions of src. *Biochim. Biophys. Acta - Rev. Cancer* **1287**, 121–149 (1996).
289. Parsons, S. J. & Parsons, J. T. Src family kinases, key regulators of signal transduction. *Oncogene* **23**, 7906–7909 (2004).
290. Creedon, H. & Brunton, V. G. . Src Kinase Inhibitors: Promising Cancer Therapeutics? *Crit. Rev. Oncog.* **17**, 145–159 (2012).
291. van Ham, T. J. *et al.* Identification of MOAG-4/SERF as a Regulator of Age-Related Proteotoxicity. *Cell* **142**, 601–612 (2010).
292. Falsone, S. F. *et al.* SERF Protein Is a Direct Modifier of Amyloid Fiber Assembly. *Cell Rep.* **2**, 358–371 (2012).

293. Wang, C., Bradley, P. & Baker, D. Protein–Protein Docking with Backbone Flexibility. *J. Mol. Biol.* **373**, 503–519 (2007).
294. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
295. Chiti, F. & Dobson, C. M. 2018 Alzheimer's disease facts and figures. *Alzheimer's Dement.* **14**, 367–429 (2018).
296. Gershenson, A. & Gierasch, L. M. Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.* **21**, 32–41 (2011).
297. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
298. Forman-Kay, J. D. & Mittag, T. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure* **21**, 1492–1499 (2013).
299. Tanaka, M. & Komi, Y. Layers of structure and function in protein aggregation. *Nat. Chem. Biol.* **11**, 373–377 (2015).
300. Ruggeri, F. S. *et al.* Influence of the β-Sheet Content on the Mechanical Properties of Aggregates during Amyloid Fibrillization. *Angew. Chemie Int. Ed.* **54**, 2462–2466 (2015).
301. Zheng, W., Tsai, M.-Y., Chen, M. & Wolynes, P. G. Exploring the aggregation free energy landscape of the amyloid-β protein (1-40). *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11835–11840 (2016).
302. Breydo, L. & Uversky, V. N. Structural, morphological, and functional diversity of amyloid oligomers. *FEBS Lett.* **589**, 2640–2648 (2015).
303. Funderburk, S. F., Marcellino, B. K. & Yue, Z. Cell "Self-Eating" (Autophagy) Mechanism in Alzheimer's Disease. *Mt. Sinai J. Med. A J. Transl. Pers. Med.* **77**, 59–68 (2010).
304. Nekrasov, E. D. *et al.* Manifestation of Huntington's disease pathology in human induced pluripotent stem cell-derived neurons. *Mol. Neurodegener.* **11**, 27 (2016).
305. Deas, E. *et al.* Alpha-Synuclein Oligomers Interact with Metal Ions to Induce Oxidative Stress and Neuronal Death in Parkinson's Disease. *Antioxid. Redox Signal.* **24**, 376–391 (2016).
306. Mahul-Mellier, A.-L. *et al.* Fibril growth and seeding capacity play key roles in α-synuclein-mediated apoptotic cell death. *Cell Death Differ.* **22**, 2107–2122 (2015).
307. Stefani, M. Biochemical and biophysical features of both oligomer/fibril and cell membrane in amyloid cytotoxicity. *FEBS J.* **277**, 4602–4613 (2010).
308. Engel, M. F. M. Membrane permeabilization by Islet Amyloid Polypeptide. *Chem. Phys. Lipids* **160**, 1–10 (2009).
309. Westermark, P., Andersson, A. & Westermark, G. T. Islet Amyloid Polypeptide, Islet Amyloid, and Diabetes Mellitus. *Physiol. Rev.* **91**, 795–826 (2011).
310. Fernández, M. S. Human IAPP amyloidogenic properties and pancreatic β-cell death. *Cell Calcium* **56**, 416–427 (2014).
311. Mukherjee, A., Morales-Scheihing, D., Butler, P. C. & Soto, C. Type 2 diabetes as a protein misfolding disease. *Trends Mol. Med.* **21**, 439–449 (2015).
312. Tomita, T. Apoptosis in pancreatic β-islet cells in Type 2 diabetes. *Bosn. J. basic Med. Sci.* **16**, 162–79 (2016).
313. Eisele, Y. S. *et al.* Targeting protein aggregation for the treatment of degenerative diseases. *Nat. Rev. Drug Discov.* **14**, 759–780 (2015).
314. Martel, A. *et al.* Membrane Permeation versus Amyloidogenicity: A Multitechnique Study

of Islet Amyloid Polypeptide Interaction with Model Membranes. *J. Am. Chem. Soc.* **139**, 137–148 (2017).
315. Young, L. M., Cao, P., Raleigh, D. P., Ashcroft, A. E. & Radford, S. E. Ion Mobility Spectrometry–Mass Spectrometry Defines the Oligomeric Intermediates in Amylin Amyloid Formation and the Mode of Action of Inhibitors. *J. Am. Chem. Soc.* **136**, 660–670 (2014).
316. Zhang, M. *et al.* Interfacial interaction and lateral association of cross-seeding assemblies between hIAPP and rIAPP oligomers. *Phys. Chem. Chem. Phys.* **17**, 10373–10382 (2015).
317. Ikeda, K. & Nakano, M. Self-Reproduction of Nanoparticles through Synergistic Self-Assembly. *Langmuir* **31**, 17–21 (2015).
318. Kelkar, D. A. & Chattopadhyay, A. The gramicidin ion channel: A model membrane protein. *Biochim. Biophys. Acta - Biomembr.* **1768**, 2011–2025 (2007).
319. mMass - Open Source Mass Spectrometry Tool. Available at: http://www.mmass.org/. (Accessed: 30th September 2019)
320. Nath, A., Miranker, A. D. & Rhoades, E. A Membrane-Bound Antiparallel Dimer of Rat Islet Amyloid Polypeptide. *Angew. Chemie Int. Ed.* **50**, 10859–10862 (2011).
321. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
322. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **97**, 50–58 (2009).
323. Walker, L. R., Marzluff, E. M., Townsend, J. A., Resager, W. C. & Marty, M. T. Native Mass Spectrometry of Antimicrobial Peptides in Lipid Nanodiscs Elucidates Complex Assembly. *Anal. Chem.* **91**, 9284–9291 (2019).
324. Li, J., Richards, M. R., Kitova, E. N. & Klassen, J. S. Delivering Transmembrane Peptide Complexes to the Gas Phase Using Nanodiscs and Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **28**, 2054–2065 (2017).